1-1-2014

Development of a Brief Screening Measure for Depression and Problem Drinking

Chris Wera
University of Denver

Development of a Brief Screening Measure for Depression and Problem Drinking

---

A Dissertation

Presented to

the Faculty of the Morgridge College of Education

University of Denver

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

Christopher Wera

November 2014

Advisor: Dr. Kathy E. Green

Author: Christopher Wera
Title: Development of a Brief Screening Measure for Depression and Problem Drinking
Advisor: Dr. Kathy E. Green
Degree Date: November 2014

ABSTRACT

This study examined the psychometric characteristics of a brief assessment measure for screening for depression and alcohol use on a college campus prior to primary care medical office visits. The measure was adapted from two widely used measures: the Patient Health Questionnaire (PHQ-4) and the Alcohol Use Disorder Identification Test (AUDIT-C). Impulsivity, which has been associated with both depression and problematic alcohol use, was also examined through additional questions. The research study investigated the psychometric properties of the PHQ-4 and the AUDIT-C, and explored if eight impulsivity items from the UPPS-P measure could enhance screening for depression and problematic alcohol consumption.

A 15-item measure was piloted with 491 college-aged individuals. The measure was examined using several analytic techniques. Exploratory factor analysis identified three factors indicating the measure contained depression (PHQ-4), alcohol use (AUDIT-C), and impulsivity factors. Rasch analysis resulted in identifying 15-item measure as multidimensional. Further Rasch analysis showed the PHQ-4, the AUDIT-C, and the impulsivity questions as unidimensional. The PHQ-4 measure showed adequate fit, scale use, and targeting for this population. Rasch analysis resulted in four-items from the eight impulsivity questions that could be treated as a scale. However, the Rasch analysis of AUDIT-C showed poor item fit and significant differential item functioning and was determined to inadequate as a scale, and so, individual items were used in subsequent analyses.

Hierarchical regression revealed a significant contribution of the impulsivity measure in explaining variance for the PHQ-4, but was lacking in explaining additional measure variance when used with the AUDIT-C individual items. Latent class analysis identified three classes, with the most interesting being male, young, and white that frequently binge drinks regularly (22% of the population).

While the 15-item scale was unsuccessful in improving identification of problematic drinking, the impulsivity items could be useful in helping to better identify depression among this population. The results also questioned the effectiveness of the AUDIT-C in screening for excessive alcohol consumption.

Further research should focus on the development of better brief screening tools in primary practice that are psychometrically sound and contain items that are not only diagnostic in nature. Inclusion of items in these instruments that explore related facets, such as impulsivity, should be explored in future development.

Keywords: Alcohol use, binge drinking, impulsivity, depression, AUDIT, PHQ-9, PHQ-4, UPPS-P, Rasch analysis

ACKNOWLEDGMENTS

I want to express my deepest thanks and eternal gratitude to my advisor, instructor, and dissertation director Dr. Kathy Green. I can't begin to express how much her incredible wisdom, even nature, encouragement, and counsel have meant. When I was at the most difficult time of my academic career, wavering on whether to continue in the pursuit of my degree—it was one word from her that guided me on.

I thank my family and friends for the support they have provided to me during the pursuit of my degree. They have given up so very much, and have stepped in at a moment's notice to help me push through even the most difficult of challenges. They mean the world to me.

But mostly . . . this is for Alison.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE: INTRODUCTION AND LITERATURE REVIEW

Overview

College years are a time when the lives of most young adults change immensely. For the majority of students, college is a time of tremendous personal and emotional growth; however, it is also a period when students experience and engage in higher levels of harmful behaviors. During their college years, students may experience high levels of depression, sometimes leading to suicidal thoughts and attempts, as well as the more common issue of excessive alcohol experimentation, misuse, and abuse. In a recent, large-scale assessment by the American College Health Association (2013) of university students that surveyed 153 institutions and over 123,000 respondents, 65.5% of the respondents had consumed alcohol within the past thirty days. Of the students that drank, nearly 32.7% had consumed five or more drinks more than once in the past two weeks and, in this same sample, students reported that, within the past twelve months, things were hopeless (45%), felt so depressed that it was difficult to function (31.3%), seriously considered suicide (7.4%), and reported attempting suicide 1.5% (American College Health Association, 2013).

Alcohol abuse, specifically binge drinking, occurs on college campuses at an alarming rate. Even without depression, binge drinking among college-aged students has reached a critical point, with a national study reporting that over 50% of all college students binge drink (i.e., for males, drinking five or more drinks; for females, drinking four or more drinks) with the purpose of getting drunk in a single evening at least once every two weeks (National Institute on Alcohol Abuse and Alcoholism [NIAAA], n.d.).

Depression and alcohol abuse have been shown to be interconnected (Gonzalez & Hewell, 2012). At the intersection of these two related concerns is the concept of impulsivity. Over the past few decades, extensive study has shown impulsivity to be a multidimensional and multifaceted construct. Depending on the instrument used, studies have shown anywhere between two and five defined facets of impulsivity. The two facets most related to college-aged depression and alcohol use appear to be perception of control over one’s actions and lack of care for negative consequences. Research shows that these facets correlate with suicidal thoughts and attempts, and also with alcohol misuse and abuse (Coskunpinar, Dir, & Cyders, 2013).

Multiple studies have shown that early intervention, especially brief interventions, can have a significant impact on drinking behavior, and may have an impact on reducing suicidal behavior. However, using screening tools for these risk factors to engage in early intervention for depression and alcohol abuse is rarely done, unless an individual experiences a crisis. A barrier to using existing screening tools is that they tend to be long, time-consuming measures that attempt to capture a wide variety of potential diagnoses. Such screening tools are also designed for longer-term therapy, rather than for brief interventions. Therefore, a psychometrically sound, short screening tool that is useful in assessing individuals for more extensive assessment and/or referral is needed.

Purpose of the Study

While there are multiple diagnostic and screening instruments for depression, alcohol use, and impulsivity, there currently is no brief screening tool that captures depression and alcohol misuse (specifically binge drinking) in one brief instrument. There is also no brief screening instrument that explores possible components of impulsivity that are related to depression and problem drinking that would assist with early identification and intervention. The purpose of this research was to investigate the effectiveness of a newly developed, fifteen-question intake screening tool for use in a primary care medical setting. The new screening tool combines items from two established and validated measures that are used to screen for depression: the PHQ-4 (Löwe et al., 2010) and the AUDIT-C for alcohol abuse (Bush, Kivlahan, McDonell, Fihn, & Bradley, 1998). Additional questions were included to enhance the alcohol abuse screening questions in order to differentiate normative college drinking, which is typically matured from or grown out of from higher risk, problem drinking, and drinking that continues well past college years. The screening tool was also intended to investigate if a targeted number of impulsivity questions can help with the early identification of dangerous levels of these harmful behaviors. Analyses were performed to investigate the latent factor structures, dimensionality, and validity of the new screening tool. The screening tool was intended to be comprised of either two or three distinct constructs—depression and problematic drinking, impulsivity, and possibly anxiety—with each construct having multiple dimensions. Methods used in instrument evaluation were the following: exploratory (EFA), item response theory (IRT), hierarchical regression (HR), and latent class analysis (LCA). The study used these analytic techniques to evaluate and refine the newly developed scale.

As part of this research study, the latent factor structures of the PHQ-4 and AUDIT-C scales were evaluated. The study aimed to leverage the complementary nature of the analyses in order to strengthen the rigor and sophistication of evaluating the newly developed screening scale.

Research Questions

- Exploratory factor analysis was used to determine if the scale held a unidimensional structure. It was hypothesized that the structure would be multidimensional.
- Rasch analysis of the full scale and potential subscales was examined to determine if there was a dimensionally appropriate structure, appropriate scale use, and the presence of differential item functioning by gender, year in college, and ethnicity—as previous research shows differences in behaviors based on these three variables—in screening instruments for depression and alcohol abuse and misuse.
- Rasch analysis of the PHQ-4 examined the dimensional structure in a college-aged population.
- Rasch analysis of the AUDIT-C examined the dimensional structure in a college-aged population.
- Hierarchical linear regression was used to test models in order to investigate the relationship between the PHQ-4, AUDIT-C, binge drinking items, and impulsivity questions.
- Latent class analysis was used to determine the presence of any undetermined classes.

All of these questions examined the psychometric properties of acceptable fit, construct reliability, and construct validity.

Review of the Literature

The Nature of Alcohol Misuse in College Settings

As the third leading cause of mortality in the US, population estimates of alcohol use suggests that over one-third of North Americans drink excessively. These rates are higher for people treated in primary care settings. Alcohol abuse, specifically binge drinking on college campuses, is rising at an alarming rate. Alcohol abuse is also a contributing factor in suicide, as 33% of decedents tested positive for alcohol (Centers for Disease Control and Prevention [CDC], 2008). Numerous studies have shown that alcohol abuse increases the risk of attempting suicide, with isolated, abusive drinking being a key indicator for depression and suicidal ideation (Gonzalez & Hewell, 2012). Further, students who have reported suicidal ideation are more likely to engage in risky behaviors (Barrios, Everett, Simon, & Brener, 2000).

While it is almost universally accepted that young adults have higher than average rates of drinking, alcohol abuse, and (specifically) binge drinking on college campuses, alcohol abuse is an increasing problem. According to the National Institute on Alcohol Abuse and Alcoholism (n.d.), about 80% of all college students drink alcohol and, of the students who drink, 50% of students in the 18 to 24 year age range engage in moderate to low risk drinking behavior. Moderate- to low-risk drinking is defined as no more than four drinks in a single day, combined with no more than 14 drinks per week for men and no more than three drinks in a single day combined with no more than seven drinks per week for women. Heavy or at-risk drinking is defined as anything exceeding the moderate- to low-risk drinking levels. Heavy or at-risk drinking is seen in college at an increasing rate, and is especially problematic, since about 25% of individuals that fall into this category have an alcohol dependency or abuse problem. Heavy drinking is often characterized by episodes of binge drinking, and binge drinking is increasing rapidly in this age group. Binge drinking is defined by the NIAAA as drinking enough in a two-hour period to have blood alcohol concentration levels reach .08g/dL. For men, this occurs after approximately five drinks and, for women, this occurs after about four drinks. In the most recent administration of the National College Health Assessment (NCHA; 2013), 32.6% of students that drank had a blood alcohol concentration of .08g/dL or higher.

A recent study among students using primary care at a student health clinic indicated that 57% of students seen are at-risk drinkers and 33% of this population met criteria for alcohol abuse (Zakletskaia, Wilson, & Fleming, 2011). Males under the age of 24 who smoke and drink at bars, or fraternities or sororities, reported episodic drinking of five or more times in the past thirty days, which is nearly double the rate of their female counterparts. For students that drink, about 25% reported that their drinking had academic consequences, such as missed classes, falling behind, performing poorly, or having lower grades due to drinking. Of even greater concern is that a large number of students developed life-long alcohol related health problems.

Some of the reasons that drinking levels increase among college students include the lax enforcement of drinking laws, limited interactions with parents or other adults, and the widespread availability of alcohol. The first few months of the first year at college are an especially risky time because of student expectations and social pressures. Alcohol consumption is particularly high for students attending schools with strong Greek systems, and students who live in fraternities and sororities drink at especially high levels. Although the majority of students tend to mature out of this period of heavy and binge drinking after college, many do not, and continue this drinking pattern. About 25% of heavy drinkers developed long-term alcohol problems (NIAAA, n.d.).

Efficacy of Interventions for Excessive Drinking

In general, students are willing to be forthcoming about their alcohol consumption and excessing drinking behaviors (Elliott, Carey, & Bolles, 2008). This suggests that college students may be open to discussion about their alcohol use, and primary healthcare clinics on college campuses are a potential source for providing alcohol interventions. The use of brief screening for alcohol as a method for early identification of problematic drinking is an important reason that measures should be psychometrically sound.

Patients are extremely responsive to brief interventions for excessive drinking by primary health providers. Several studies have shown that a five-minute intervention with advice from a primary care provider can reduce alcohol consumption by 25% (Kaner, Heather, Brodie, Lock, & McAvoy, 2001). Preventive alcohol screening programs have a fairly low refusal rate, and the impact on drinking behaviors through the use of simple screenings can make a significant impact on patients' lives.

A meta-analysis by Beich, Thorsen, and Rollnick (2003) to investigate the effectiveness of screenings prior to behavioral interventions in general practice settings found that while a number of interventions were evaluated, the outcome was that roughly 10% of patients decreased their drinking due to the behavioral intervention. Several other systematic reviews and meta-analysis studies showed similar reductions in drinking, ranging from 13% to 34%, based on the method of screening and the frequency and length of the behavioral intervention (Whitlock et al., 2004). Another summary of a systematic review and meta-analysis from an international base of studies supports the use of brief interventions (Kahan, Wilson, & Becker, 1995). Using time spans of six to twelve months, reductions from heavy to acceptable rates of alcohol consumption ranged from 10% to 19%. The behavioral intervention was also effective with binge drinkers (Hyman, 2006). A more recent meta-analysis that included 23 trials showed that brief interventions (i.e., ten to fifteen minutes) were the most effective in reducing problem drinking, compared to very brief (i.e., five-minute interventions) interventions (Jonas et al., 2012). In this summary, consumption decreased from the baseline by a weighted mean difference of 3.6 drinks/week [95% CI, 2.4 to 4.8 drinks/week], with 12% fewer adults reporting episodes of heavy drinking episodes (Jonas et al., 2012).

The effect of behavioral interventions is also seen in college populations. Kulesza, Apperson, Larimer, and Copeland (2010) investigated using different intervention durations to evaluate if longer interventions had a different effect on drinking reduction outcomes. While the sample size was small, the study showed that a brief intervention (i.e., a ten-minute intervention) was significantly more effective in reducing drinking rates than none at all. However, there was no significant difference when a 50-minute intervention was used, compared to no intervention (control group). This indicates that a brief intervention is as effective in reducing alcohol use and related problems in high-risk drinkers as is the typical counseling model.

Depression

Another major issue in a college-aged population is the prevalence of depression and depressive symptoms, especially when combined with excessive alcohol consumption. An estimated one in ten U.S. adults report being diagnosed with clinical depression; although, the number of adults with undiagnosed depression is thought to be several times higher than this, as most adults experience some sort of crisis before being diagnosed by a healthcare professional (CDC, 2011). Undiagnosed depression is thought to be higher among young adults, and especially prevalent among college students. According to the American College Health Association (2013), as of spring 2013, 10.9% of college students had been diagnosed or treated for depression in the past year. Again, these statistics capture only the number of students that have been diagnosed by a healthcare professional. It is estimated a large percentage of this population has depression-related symptoms that go unidentified and are untreated.

Effectiveness of Assessment and Interventions for Depression

One of the most effective ways to address depression is to provide early intervention (Pyne et al., 2003). With intervention, mental health professionals can provide simple, effective treatments (such as short-term counseling), rather than running the risk of having underlying symptoms worsen if left untreated. In an effort to identify and intervene early to address depressive symptoms, recent efforts have focused on using screening tools for patients at or during primary care medical visits. Published reports have associated nearly 70% of adult primary care visits as having an underlying mental health condition that is often depression-related (Strosahl, 1996). Since most individuals visit a primary care provider at least once per year, these screening tools provide an opportunity for a brief mental health assessment.

Depending on the model, performing a depression intervention after a screening can help the individual better understand his or her symptoms. It is becoming more common to find larger medical clinics that have a mental health counselor on hand (Oopik, Alouja, Kalda, & Maaroos, 2006). Having an in-house mental health provider available is an extremely effective method of improving mental health treatment. Even when this option is not available, many times just speaking to the provider about issues that would have gone undetected without a screening provides the opportunity for the provider and the patient to discuss treatment options. A recent study showed that having a medical team trained in mental health interventions improved depressive symptoms, physical functioning, and satisfaction with care (Rost, Nutting, Smith, Elliott, & Dickinson, 2002).

The Role of Impulsivity in Alcohol Use

Impulsivity is a complex and important construct that, historically, has been approached from a multitude of perspectives, depending on theoretical orientation. The literature on this multidimensional construct is as varied as the instruments that attempt to measure it. The construct of impulsivity can be used to describe normal behaviors as well as clinically defined personality disorders. One of the more common definitions of impulsiveness is “a predisposition toward rapid, unplanned reactions to internal or external stimuli without regard to the negative consequences of these reactions to the impulsive individuals or others” (Moeller, Barratt, Dougherty, Schmitz, & Swann, 2001, p. 1785. The fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 2013) describes impulsivity as possibly the most common diagnosis after subjective distress, and dedicates an entire section to impulse control disorders. However, there is no general consensus on a comprehensive, theoretical framework about the components of impulsivity.

A body of research has focused on behavioral tasks for assessing impulsivity (e.g., Stop-Go Task, Mirror-Tracing Persistence Task, and Balloon Analogue Risk Task) (Arce & Santisteban, 2006). For the most part, these tasks focus on inhibitory control, persistence, risk-taking, and delayed discounting related to individual impulsivity. Since the focus of this dissertation is to identify a brief screening model for identifying selected facets of impulsivity, these instruments are beyond the scope of this paper; however, given time for additional screening, it is possible to consider them as additional diagnostic tools. There has also been an immense amount of research on impulsivity related to the neurochemistry of impulsivity, the development of scales to measure impulsivity levels, and to define the facets that make up this broad construct (Moeller, Barratt, Dougherty, Schmitz, & Swann, 2001; Miller, 2004; Lejuez, 2010).

The inability of researchers to come to a common definition of impulsivity has led to extensive disagreement regarding study design and outcomes; however, it has also focused a broad and complex body of research on this topic. Research on impulsivity has mainly focused on the negative outcomes related to psychopathology, including aggression, poor decision-making, attention deficient disorder (ADD), and alcohol and substance abuse; it has also focused on more mundane issues, such as learning and workplace behaviors.

Impulsivity and Alcohol Abuse

Related to emerging adults, research has consistently indicated an association between impulsivity and alcohol use. While there is a large body of work that implies a relationship between impulsivity and alcohol use disorders (AUD), until the creation of the UPPS (Urgency, Premeditation (lack of), Perseverance (lack of), and Sensation

Seeking) and UPPS-P (Urgency (negative), Premeditation (lack of), Perseverance (lack of), Sensation Seeking, and Urgency (positive)) instruments, there was very little research that examined the distinct facets that make up impulsivity and how they are associated with binge drinking and AUD. The UPPS and its variants has become the baseline assessment tool to investigate these relationships. Recent research has found that urgency (both negative and positive) and sensation seeking have been related to binge drinking and AUD. Current studies have also suggested that these distinct traits may have a role in the escalation of alcohol use and the development of AUDs during emerging adulthood. The results of this research are relevant, as it may help identify problematic drinking in a high-risk population, such as young adults. As stated before, rates of alcohol consumption in this population range from 50% for ages 18-20, to 70% in the 21-25 aged population (Substance Abuse and Mental Health Services [SAMHSA], 2008). Binge drinking rates in this age population regularly exceeds 50%.

Previous studies have found that urgency was related to AUDs, and sensation-seeking was also associated with frequency of drinking. Research by Shin, Hong, and Jeon (2012) utilized the UPPS to assess drinking frequency and amounts of the past twelve months, and found that urgency and sensation-seeking were positively associated with frequency of alcohol use. In this study, the Poisson regression model that predicts alcohol-related problems from impulsivity factors was statistically significant ($\chi 2\ 9 = 99.8$, $p<.001$), with higher scores on urgency and sensation-seeking associated with greater alcohol problems. This study also indicated that urgency is most strongly associated with AUDs, and sensation-seeking was strongly related to binge drinking. Coskunipinar et al. (2013) conducted a meta-analysis to examine the varied relationship sizes of 96 studies, and summarized the association between impulsivity facets and AUDs. Consistent with prior research, the results indicated that impulsivity and alcohol are related (r = 0.28), although effect sizes varied significantly across the studies (-0.05 to 1.02). The meta-analysis revealed that lack of perseverance was the best indicator of drinking quantity (r = 0.32), whereas all five traits were linked to drinking frequency. Drinking problems were best indicated by negative (r = .05) and positive (r = 0.34) urgency, with alcohol dependence indicated by negative urgency (r = 0.38) and lack of planning (r = 0.37). These findings support the body of research and literature that shows that specific impulsivity facets relate differently to alcohol consumption patterns.

The Role of Impulsivity in Depression

Earlier research on impulsivity and depression showed a clear relationship between the two, with assessment instruments for each showing the correlation. For example, one study demonstrated that impulsivity levels were higher among individuals with suicidal attempts and depression, as compared to those individual with non-suicidal depression (Corruble, Darny, & Guelfi, 1999).

Research shows that impulsive individuals consume more alcohol, while suicidal behavior and depression have long been recognized as a problem in alcohol-dependent people. Individuals that score higher on an impulsivity measure also tend to score higher on measures of depression (Koller, Preuss, Bottlender, Wenzel, & Soyka, 2002). When alcohol misuse—including binge drinking, drinking to cope, and AUDs—are added to depressive symptoms, the likelihood of suicide attempts or non-suicidal self-injury increases. Gonzalez and Hewell (2012) showed that suicidal ideation accounted for the most variance in drinking to cope and, further, the impulsivity factor of negative urgency was significantly associated with drinking to cope.

Gonzalez and Hewell (2012) administered the AUDIT, the UPPS, and other depression measures, and demonstrated a significant interaction between urgency, alcohol use, and depressive symptoms. This research supported the linkage between suicide risk factors and depressive symptoms, and confirmed a direct association between problematic drinking and suicidal ideation. This finding suggests that addressing drinking to cope in at-risk individuals through assessing underlying impulsivity factors may aid in addressing depression, and may also reduce the likelihood of suicide attempts.

Brief Screening Tools

The literature shows that depression, alcohol misuse, and impulsivity are all complex constructs. Each of these constructs assesses multiple diagnoses that can be comorbid, and can involve other medical and mental health issues. Over the past three decades, there has been an emphasis on developing shorter, more targeted screening instruments for use in opportunistic settings, including emergency departments or primary medical care offices, rather than solely in a traditional counseling setting. The challenge of providing early intervention is to use brief and convenient screening opportunities to bring awareness of underlying issues to the individual and the healthcare provider.

One challenge of performing brief screenings is that most available instruments are designed for diagnosis, rather than for use in a brief screening setting. Longer diagnostic instruments are designed to address many facets of each mental health construct. The different measures available for depression is one example: There are dozens of different, validated instruments that are designed for different diagnoses of depression, for different ages and genders, and for different severities of this mental illness. It is not uncommon to find even simple screenings for depression to contain over fifty items, in an attempt to provide clinicians with enough information to begin providing an appropriate treatment. Currently, the most widely used tool for impulsivity contains nearly sixty items.

The purpose of using brief screenings in a primary healthcare setting is to reveal issues that the patient may not be have intended to discuss during the visit, but which may be relevant to the presenting problem. Because of the relationship between the provider and the patient, there is an opportunity to discuss issues revealed by a screening instrument confidentially, while the patient may be more open to change.

Brief screenings tools tend to be subsets of longer, diagnostic instruments, and are typically focused on one issue, such as depression or alcohol consumption. Among other issues, instrument choice depends on practicality, logistics, and specifics of the target population, as well as setting, and resource allocation. The key with brief screening instruments is that they must be short, reveal the most common issues, and provide the clinician with a baseline for further care, referral, and diagnosis. Ease of use and scoring are as important as validity and reliability so that a screening instrument can be useful in most primary medical care settings. Screening measures must be brief and raise suspicions or detect a potential problem. They are not designed to confirm a diagnosis. A positive screening should be evaluated further. Shorter screenings enhance the feasibility of use in primary and urgent care settings.

Depending on the situation where the screening occurs, research suggests that a one- or two-question instrument could be an effective brief screening tool. Smith, Schmidt, Allensworth-Davies, and Saitz (2009) found that simply asking how many times in the past year have the patient has engaged in hazardous or binge drinking was 81.8% sensitive (95% CI, 73.1% to 84.4%) and 79.3% specific (95% CI, 73.1% to 84.4%) for unhealthy alcohol use. These results are similar to longer versions of alcohol screenings, such as the AUDIT.

One challenge in using too brief a screening instrument is that if positive, a secondary screen is typically employed. This takes up further provider time, and gives the patient an opportunity to change his or her answers during the visit.

Results consistently demonstrated that a brief intervention is more effective than no intervention. Given that resources are often limited, electronic screening instruments have become increasingly popular. A recent survey of alcohol consumption, electronic screening, and brief interventions showed mixed results (Bewick et al., 2008). However, an earlier study indicated that web-based interventions could attract many users who would not have otherwise availed themselves of this resource (Saitz et al., 2004). The results of this study indicated that users with alcohol dependency were more likely to use electronic resources than drinkers in the hazardous drinking category.

Measures of Study Constructs

PHQ-4 Depression Measure

While brief screening instruments are not the norm in traditional counseling or medical offices, there are a number of screening instruments that have been developed for use specifically in primary care settings. One of the most frequently used of these medical primary care instruments is the Patient Health Questionnaire-9 (PHQ-9), a self-administered, ten-item questionnaire based on the PRIME-MD Patient Health Questionnaire diagnostic survey ["Patient Health," n.d.]. The purpose of the PHQ-9 is to facilitate the recognition of symptoms for the most common mental health depressive disorders. Eight questions on the PHQ-9 score responses to symptoms over the past two weeks in one area of the eight DSM-IV criteria for depression. There is an additional question on suicidal ideation and a final, non-scored question that is used to ascertain the overall impact of any of the other nine questions on overall functioning (Kroenke, Spitzer, & Williams, 2001). The response scale for the nine scored questions is as follows: “0” for not at all; “1” for several days; “2” for more than half of the days; and “3” for nearly every day. The total scores for each of the nine diagnostic questions are summed. Score cut-offs are used to create categories that were found to correlate with different clinical depression diagnoses. The score cut-offs delineate minimal depression (1-4), mild depression (5-9), moderate depression (10-14), moderately severe depression (15-19), and severe depression (20-27). The PHQ-9 can be completed in an average of two minutes, it is simple to score, and medical staff can interpret results with minimal training. A major advantage of the PHQ-9 is that it is a dual-purpose diagnostic instrument. It is valuable not only for assisting with diagnosis, but also for assessing the severity of symptoms based on the value of the score.

The PHQ-9 was originally developed and validated in 2001, with underwriting from an education grant from Pfizer U.S. Pharmaceuticals. The PHQ-9 is widely used, since it is free and no permission is needed to use or reproduce it. This instrument has been extensively used by the United States Veterans Administration during routine primary care medical visits, and also by several, large health maintenance organizations. In these organizations, use of the PHQ-9 has served to increase early mental health intervention, improve the recognition of depressive symptoms, and has resulted in large cost savings (Pyne et al., 2003).

The initial study of the PHQ-9 was published in 2001 to examine its reliability, efficiency, and operating characteristics as a diagnostic depression instrument, and also to verify construct validity as a measure of depression severity (Kroenke et al., 2001). The development study examined the results of 6,000 completed surveys from both primary care and obstetrics-gynecology clinics, and compared the PHQ-9 results to another validated measure. The internal consistency reliability was high with a Cronbach's alpha of 0.89 in the primary care population, and 0.86 in the obstetrics-gynecology population. Using a sample of 580 primary care patients, criterion-related validity was demonstrated through independent re-interviews, and construct validity was established by a collation of PHQ-9 scores and functional status, disability days, and symptom-related difficulty. Generalizability of validity coefficients was established by comparing the primary care and obstetrics-gynecology samples.

While the PHQ-9 is a measurement developed relatively recently, it is thought to be one of the most widely used and validated brief screening instruments. It has widespread use in federally funded research programs, and is the standard measure of depression at the Veterans Administration facilities and in managed healthcare settings nationally and internationally (Kroenke, Spitzer, Williams, & Löwe, 2010). In 2007, its lead authors developed an even shorter four-item instrument intended to measure depression and anxiety as well as the PHQ-9. The four questions used for the newly developed PHQ-4 are the two items from the PHQ-2 (which itself is a subset of the PHQ-9) and the two questions from the GAD-2 (General Anxiety Disorder). The GAD-2, an anxiety measure created by joint authors, and is based on the GAD-7, is an extensively validated anxiety disorder measure (Löwe et al., 2008). While the GAD-2 does not have the exact wording as the anxiety questions in the PHQ-9, it is psychometrically equivalent (Kroenke et al., 2010). The PHQ-4 was examined for construct and factorial validity with other anxiety and depression scales, and was found to correlate as well with or better than other scales, and internal consistency reliability was high for all scales, with Cronbach's alpha exceeding 0.80 for all scales. One of the benefits of the PHQ-4 over the PHQ-2, which focuses on identifying and measuring depression, is the inclusion of the GAD-2 questions on anxiety. Questions on the GAD-2 are substantially better than the PHQ-2 in detecting the most common anxiety disorders, which enhances intervention and treatment. Cutoff scores for the two depression and the two anxiety items on the PHQ-4 is ≥ 3. Overall, the research showed that the PHQ-4 is an extremely efficient "ultra-brief" instrument for detecting depression and anxiety that contains two subscores to make identifying each issue possible.

While creating a measurement of anxiety is not part of this study, confirmatory factor analysis has shown that the PHQ-4 has an acceptable unidimensional fit for depression and anxiety, as well as a slightly better two-dimensional fit for depression and anxiety. In many patients with depression, there is comorbidity with anxiety in up to 50% of cases (Löwe et al., 2008), which is likely the cause of the fit to one- and two-dimensional models. Furthermore, the fit is likely due to the inclusion of questions from the GAD-2 on the PHQ-4 that are slightly different in wording from the anxiety questions on the PHQ-9, which is unidimensional.

A follow-up study on the PHQ-4 was intended to validate this instrument in the general population (Löwe et al., 2008). Using a sample of over 5,000 cases in Germany, construct validity was supported by intercorrelations with other self-report measures; the two-dimensional measure showed good fit with a RMSEA of .027; 90% CI .023- .032. PHQ-2; and GAD-2 scores of three corresponded to percentile ranks of 93.4% and 95.2%; and scores of five corresponded to ranks of 99.0% and 99.2%, respectively. For use in clinical settings, the overall score should be used as the indication of either depression or anxiety. Overall scores of six or greater (percentile 95.7%) is recommended as a "yellow flag," and scores of nine or greater (percentile 99.1%) as a "red flag" for the presence of either depression or anxiety. It is recommended that an examination of the total score should be used for initial screening, and examination of the two subscales scores—with a cutoff of three or higher on each subscale—should be used to investigate the presence of depression, anxiety, or both. Finally, the study investigated the similarities of the German population compared to the United States population, and reported that no substantial differences were present. This indicates that the results and cutoffs for an American population should be similar.

Alcohol Use Measures

Until the mid-1980s, the CAGE, a four-item brief screening instrument, was the only alcohol appraisal tool available (Bush, Kivlahan, McDonnel, Fihn, & Bradley, 1998; Meneses-Gaya et al., 2010). In the early 1990s, the Alcohol Use Disorders Identification Test (AUDIT; Babor, Higgins-Biddle, Saunders, & Monteiro, 2001) was developed, and was found to have solid psychometric properties for use among adults. The AUDIT is ten questions, and is difficult to score and use in many opportunistic settings. Since the 1990s, there has been a renewed focus on developing and validating brief screening tools that focus on identifying hazardous drinking and alcohol use disorders (AUD). Most of the brief alcohol screening instruments developed are shortened versions of the AUDIT, the CAGE, or have been varieties or combinations of both—sometimes with added questions from other widely used (though longer) screening instruments, such as the twenty-question CAPS (Maddock, Laforge, Rossi, & O'Hare, 2001).

Since the World Health Organization adopted the AUDIT as its baseline-screening instrument, extensive research has been performed on this tool. As such, shorter versions of the AUDIT are those most commonly focused on by researchers when developing even shorter instruments. Perhaps the most widely used and researched variation of the AUDIT is the AUDIT-C, a three-question instrument that has been found to have nearly identical psychometric properties as its longer predecessor (Meneses-Gaya et al., 2010). Other research shows that even asking one or two questions regarding drinking behaviors can be as effective as using instruments that utilize more items to screen for alcohol use. Such brief screenings can be most effective in an emergency department setting (Hill, Pettit, Green, Morgan, & Schatte, 2012). However, a limitation with this approach is that little information is gathered, and valuable provider time is used to ask additional questions in order to gain enough information to determine if an intervention is needed. The appropriate number of questions needed to determine if an intervention is necessary or not is a key decision in choosing a brief screening instrument, as it will determine if provider time for treatment is used effectively.

AUDIT and AUDIT-C

The AUDIT (Babor et al., 2001) is the most widely used alcohol-screening tool internationally. This screening instrument assesses levels of alcohol consumption as well as problems that result from drinking. It was designed to identify primary care patients with drinking problems who would benefit from a brief alcohol intervention. A college-aged population was not the original population selected for the survey development, and follow-up research on general populations has had mixed results on the threshold for identifying risky or alcohol use disorders. Daily drinking estimation and concurrent recall methods are commonly used to gather information.

The AUDIT-C (Bush et al., 1998) focuses on levels of alcohol consumption rather than on negative drinking consequences. This focus can be especially useful when screening for excessive levels of consumption or binge drinking among adolescents or college-aged students. Furthermore, excessive consumption has been shown to be predictive of later AUDs (Hill et al., 2000). The AUDIT-C was originally validated as a three-item screen for alcohol misuse, and was implemented nationally at Veterans Affairs clinics in the United States. In this population, a threshold score of equal to or greater than four drinks for men, and three for women, was determined as optimal for intervention. Further validation was done using European samples, with thresholds of equal to or greater than five drinks for men and women. A study by Bradley et al. (2007) indicated threshold scores of greater than or equal to four drinks for men, and three for women, simultaneously maximized sensitivity and specificity (.86 and .89 for men, and .73 and .91 for women, respectively). This study also compared the AUDIT-C to other validated and widely used measures, including the CAGE, a version of the CAGE with added questions that included consumption measures, and the full ten-question AUDIT. The results of this study demonstrated that the AUDIT-C performed as well as the ten-question AUDIT, and better than the augmented CAGE. When screening the general population for alcohol misuse in the past year, the optimum screening threshold was four or more drinks for men, and two or three for women. When screening for alcohol use disorders, the cut-off screening score is slightly higher. The optimum cut-off for men was between four and five drinks and three and four drinks for women (Bradley et al., 2007).

The research suggests that in populations where there is low prevalence of alcohol misuse, a lower threshold should be used. Alternatively, in populations with higher levels of alcohol misuse, a higher threshold is recommended.

One limitation of the AUDIT-C is that while it captures consumption, it does not focus on the negative consequences of excessive drinking. Binge drinking is associated with high rates of negative outcomes, such as hangovers, fights or arguments, unintended sexual intercourse, or self-harm. Therefore, the AUDIT-C alone may not fully capture the consequences of binge drinking, and additional questions might be needed to identify this behavior.

UPPS and UPPS-P

Over the last fifty years, impulsivity has been examined extensively with the Barratt Impulsiveness Scale (BIS), which is currently in its eleventh revision (Patton & Stanford, 1995). The BIS-11 is a thirty-item self-report instrument designed to describe the personality or behavioral construct of impulsivity. The internal consistency coefficients for the BIS-11 total scores range from 0.79 to 0.83 in different populations. The development of this scale was informed by data gathered from four diverse models:

medical, psychological, behavioral, and social. It is arguably the most widely used impulsivity scale, with 551 citations to its use as of 2009 (Stanford et al., 2009). The model underlying the BIS hypothesized that impulsiveness was a multi-dimensional construct that was related to fluctuations in the ability to make decisions. Development of the scale continued to be refined over the next several decades, with a focus to better define the factors being measured that make up impulsivity and consistency. Barratt (1959) developed the BIS-11 to more specifically define the sub-facets of impulsivity. Principle factor analysis of this instrument by Patton and Stanford (1995) produced six first-order factors and three second-order factors. The higher-order factors were defined as attentional, motor impulsiveness, and non-planning. The BIS-11 and its first- and second-order factors have been used extensively over the past twenty years to show relationships to different clinical syndromes, such as substance abuse, mood disorders, suicide attempts, and other psychological disorders. It is also used with typical populations.

More recent work has expanded beyond the BIS-11 in an attempt to better define the construct of impulsivity in relation to validated personality measures, rather than as a stand-alone construct. This research used factor analysis of general personality instruments, which resulted in a five-factor model. Whiteside and Lynam (2001) presented a new scale that resulted from examining the relationships among several commonly used instruments, including the BIS-11, to the five-factor model of personality. This new instrument breaks from past impulsivity scales in that the basis of the research was not to measure impulsivity as a score per se, but to identify the factors of personality that contribute to impulsivity—factors that can be measured individually.

This research—using principal components analysis of ten instruments, some with as many as 240 items—identified a four-factor solution that explained 66% of the variance in the measures. The four factors were labeled as follows: Premeditation, Urgency, Sensation Seeking, and (lack of) Perseverance (Whiteside & Lynam, 2001). From the initial analysis, 50 items with the highest factor loadings were selected from the examined instruments in order to measure the four factors. This number was reduced to 45 items, after removing five items that were deemed duplicative in nature. Final internal consistency coefficients were 0.91, 0.86, 0.90, and 0.82. The results for convergent, corrected item-total correlations had a mean of .58 (range 0.38- 0.79) and divergent item total correlation with a range of .17 (range 0.05 to 0.33), which suggest good convergent and divergent relation among items. The four facets identified above point to discrete processes that lead to impulsive behaviors, and are not considered variations of impulsivity. Urgency measures the tendency to give in to strong impulses when accompanied by negative emotions, such as depression. Perseverance (a lack of) is defined as the ability to persist despite boredom. Premeditation (or lack of) assesses the ability to think through the possible consequences of actions, and Sensation Seeking identifies preference for stimulation or excitement.

This newly developed scale was named the UPPS scale (Urgency, (lack of) Premeditation, (lack of) Perseverance, Sensation Seeking), and is currently the most widely studied impulsivity instrument (Whiteside & Lynam, 2001). After the release of the UPPS scale, research accelerated on the four personality facets and their relationship with impulsivity disorders. The UPPS instrument has recently been modified with a fifth pathway, Positive Urgency, based on the work of Cyders et al. (2007), which seeks to better identify the Urgency facet into Negative and Positive urgency. There is a previous body of research that focused on the concept of Positive Urgency and excessive risk-taking. Cyders et al. found that positive urgency was the only type of impulsivity that predicted risky behavior, has a significant interaction with alcohol expectations, and helped explain the variance in problem drinking. The resulting UPPS-P scale is a 59-item instrument that has been used extensively in alcohol impulsivity research in the past few years.

More recent work has been done to reduce the number of questions on the UPPS and UPPS-P instruments in order to be used in more brief screening settings. Recently, a French language, 20-item UPPS-P was developed using the highest loading items from the 59-item UPPS-P instrument (Billieux et al., 2012). Each of the five different impulsivity facets has four items. The results of this research indicated that two models fit the data: one with the five impulsivity facets, and another two-level hierarchical model with the combined urgency and lack of conscientious (lack of premeditation and lack of perseverance) as the higher-order factors. Internal consistency reliability coefficients for the French short UPPS-P were near those for the longer English version. The main benefit of this shorter version is its applicability for use outside of clinical counseling visits.

CHAPTER TWO: METHOD

This chapter explains how this study was conducted by presenting a description of the data set, sample, and variables used. Included are sections that explain the analytic methods used to test the significance of the relationships between the two validated instruments (the PHQ-4 and the AUDIT-C) with the other ten experimental questions, the study participants and how they were selected, and the instruments used to collect the data.

Many constructs in the medical and mental health areas, such as depression or a tendency for alcohol misuse and impulsivity, have a component that can be measured directly and a component that may not be observable. Researchers create measures that leverage what can be measured, such as number of drinks consumed, with other items that serve as proxies to represent the underlying phenomenon, which are known as a latent variable (DeVellis, 2003). An example of a latent variable is intelligence. While there is no direct way to measure intelligence, empirically, assessment is possible by measuring or observing variables that infer intelligence. By using a theoretical framework for intelligence, measures can be constructed so that individuals that are thought to have higher intelligence would achieve a higher score.

The purpose of this study was to develop a more brief and accurate instrument for use during a primary care visit to identify individuals who are the most at risk for depression and AUDs. Leveraging the existing research done on the PHQ-4 and AUDIT-

C with select questions on four of the identified facets of impulsivity, a benefit of this work would be to give providers better insight into potential underlying mental health issues. It was expected that the selected impulsivity questions would help identify the severity of the underlying depression and alcohol abuse facets. Psychometric analysis was performed to represent the relationships between the observed and any latent variables in the most parsimonious way. Analyses were proposed to identify and confirm a reduced set of latent variables that underlie the represented items (Gliner, Morgan, & Leech, 2009).

Analytic Methods

Exploratory Factor Analysis

Exploratory factor analysis (EFA) has traditionally been used to explore the underlying structure of a set of observed variables, without specifying a preconceived structure for the outcome (Child, 1990). EFA is a variable reduction technique that identifies the number of latent constructs and the underlying factor structure of a set of variables. It was initially developed over a century ago by Spearman, and has become one of the most widely used statistical methods (Fabrigar, Wegner, MacCallum, & Strahan, 1999). EFA hypothesizes the presence of an underlying construct, or latent trait, which is a characteristic that cannot be measured, and is often used when the researchers have no solid hypothesis about the nature of the underlying factor structure of the measure. EFA estimates factors that influence responses on observed variables, and allows descriptive identification of the number of latent constructs. EFA can be an appropriate form of analysis if the goal is to arrive at a parsimonious representation of the associations of the measured variables.

EFA is based on the common factor model that theorizes that each measured variable is a linear function of one or more common factor and one unique factor. Common factors are unobservable latent variables that influence more than one measured variable, and are presumed to account for the correlations or covariance among the measured variables. The goal of factor analysis is to help provide meaning in order to explain variation among variables by using a few newly created variables. This is achieved by estimating the pattern of relations between the common factors and each of the measured variables by examining factor loadings. Typically, eigenvalues are examined to decide on the number of factors.

The most critical methodological issue a researcher faces when determining whether or not to use EFA is what variables to include and the size of the sample. There are many different suggestions on adequate sample size for EFA, ranging from five participants per variable, but never less than 100 (Gorsuch, 1983), to a ten-to-one ratio (Nunnally & Bernstein, 1994). Recent sample size recommendations are based on the communalities of the variables and the number of variables per factor. If there are at least five variables per factor, and communalities are high (.70 or higher), the sample size can be as low as 100; however, under more moderate conditions, a sample size of 200 or more is appropriate. When working with unknown communalities, a sample size of 400 or more is recommended (MacCallum, Widaman, Preacher, & Hong, 2001).

There are three steps in conducting an EFA: (1) deciding the number of factors; (2) choosing an extraction method; and (3) choosing a rotation method. There are numerous approaches to deciding the number of factors. One is to generate a scree plot, which is a two-dimensional graph with factors on one axis and eigenvalues on the other.

Eigenvalues are produced when performing principal components analysis (PCA), and represent the variance accounted for by each underlying factor. Interpretation of the scree plot is to retain factors with eigenvalues above the plot “elbow.” Another common approach for choosing factors is to use the Kaiser-Guttman rule, which identifies factors with eigenvalues greater than 1.0 as interpretable. Another approach to determining the number of factors is to use parallel analysis (Tabachnick & Fidell, 2007). Factors are retained for interpretation if the eigenvalue from the EFA exceeds the eigenvalue from the parallel analysis of the simulated data of the same matrix size.

Once the number of factors is decided, an analysis is run to obtain loadings for each factor. There are several different extraction methods, but two of the most popular are PCA, which assumes that there is no measurement error, and maximum likelihood, which is basically canonical factoring, alpha factoring, and principal axis factoring with iterated communalities (a least squares method). Since PCA can produce poor estimates of the population loading in small samples, the best empirically supported methods are principal axis factoring and maximum likelihood approaches. Typically, when samples are large, all of the above methods have similar results.

The extraction method produces factor loadings for every item on every extracted factor. The desirable outcome is simple structure with most items having a large loading on one factor, and small loadings on the other factors. However, factor solutions rarely yield a simple structure without using a rotational technique—if multiple factors are retained. Once an initial solution is obtained, the loadings are rotated to maximize high loadings and to minimize low loadings to find the most parsimonious solution. Rotations are either orthogonal or oblique. Orthogonal rotation assumes that the factors are uncorrelated, which is rarely the research assumption. There are three common methods of orthogonal rotation: varimax, quartimax, and equimax. Oblique rotation derives factor loadings based on the assumption that the factors are correlated, which is probably the case for many measures. In addition to the loadings, oblique rotation gives the correlation between the factors. The most common methods of oblique rotation are oblimin, promax, and direct quartimin. Since oblique rotation provides estimates of the correlations among common factors, oblique rotation is thought to provide more information in order to produce a more accurate representation of how the constructs are likely related.

Assumptions of EFA are interval or ratio level of measurement, the relationship between the observed variables is linear and that there is a similar, preferably normal, distribution for each observed variable, and multivariate normality.

In the current study, EFA was used to identify the latent constructs underlying the set of fifteen items. Specifically, parallel analysis was used to determine the appropriate number of factors to retain. If more than one factor was indicated, the extraction technique used was principal axis factoring with oblique rotation. Items were considered to load adequately on a factor if the loading exceeded .30; items were considered to crossload if loadings on two or more factors differed by less than .10.

Rasch Analysis

For scale evaluation and calibration, item response theory—specifically, the Rasch model—was used. The underlying theory of the Rasch model is that it seeks to determine how well the scale works as an unbiased measure with items arranged in a monotonically increasing pattern by item position or difficulty. The Rasch model can be used with either dichotomous or polytomous response scales. For polytomous scales, Rasch analysis displays the response scale structure and fit as well as item and person fit, and provides estimates of the difficulty of each response or scale step for each item. Rasch model fit is tested by a series of fit indices, such as the mean square fit with an expectation of a fit of 1.0, with a range between 0.0 and infinity. Fit mean square is modeled to be 1.0 when the data fit the model. Values greater than 1.0 indicate underfit, and values less than 1.0 indicating overfit. Underfit indicates excessive noise in the data, while overfit indicates possible overlapping item content.

The Rasch model assumes unidimensionality. Unidimensionality means that the items forming the instrument all measure the same singular variable—the latent construct—which, in this instrument, is depression, alcohol use, or impulsivity. The evaluation of dimensionality is necessary to support the evidence of validity in an item response theory framework. Having a unidimensional structure allows for examination of item and ability without bias (Yu, Popp, DiGangi, & Jannasch-Pennell, 2007). Identifying poorly fitting items, and removing them from the scale, improves the unidimensionality of the instrument.

An assumption of the Rasch model is that items in the scale measure a single latent construct. The PHQ-4 was developed to screen for a multitude of aspects of the latent construct of depression. The AUDIT-Cmeasures the latent construct of alcohol misuse, and questions were examined for unidimensionality of an impulsivity construct. The 15 items comprising the proposal measure were also examined for dimensional structure, where it was anticipated that three dimensions would be identified. Dimensionality was examined in this study, using: (a) overall fit of the data to a one-dimensional model; (b) a Rasch principal components analysis of residuals; and (c) individual item fit.

Fit is measured in two ways in the Rasch model. Infit, a measure weighed by the distance between person and item location, and outfit, which is an unweighted measure. Fit statistics are transformations of chi-square statistics. It was expected that the mean square (MNSQ) indices would be 1.0 if the data fit the model exactly. Acceptable infit and outfit values for items with a Likert scale response fall within a range of 0.6-1.4 (Walker, Engelhard, & Thompson, 2012). Values above 1.4 indicate that the data contain more variability than was expected based on the model, whereas values below 0.6 indicated less variability than expected based on the model. Person fit in the Rasch model is an indication of whether individuals respond in a consistent manner. Poor person fit indicates inconsistent or erratic responses. Item fit relates to the functioning of the items. Good item fit indicates that the questions are logical, form a continuum, and are related to a single construct. Poor item fit indicates that the item may be too complex or difficult in relation to the rest of the scale, or may not be consistent in measuring the construct being examined. The average MNSQ for the calibration sample for infit was 1.10 and .89 for outfit, indicating good fit. Fit indices may also be transformed to a standardized metric, with an expected value of 0.0. In this study, the mean square fit indices were used.

Invariance is critical to the usefulness of the PHQ-4 in screening college-aged individuals for depression and alcohol misuse. Invariance is defined as consistency in the ordering of item responses across person groups with differing characteristics. Examples of invariant measures include height, weight, and temperature. Measures of health, such as pain or depression, tend to be historically fairly invariant (Löwe et al., 2010). Failure of invariance means that it is impossible to compare samples because the construct is construed differently by each sample. Bond and Fox (2001) define invariance as maintenance of identity in the meaning of a variable from one time or group to the next.

Invariance is assessed with differential item functioning, which is based on a shift in meaning expressed as item location over time or between groups. Differential item functioning (DIF) is calculated by generating the logit position by group or time, and dividing the difference by the combined standard error. If DIF is evidenced in the items, it indicates that identity in meaning was violated (Stark, Chernyshenko, Chuah, Lee, & Wadlington, 2001). DIF may occur due to construct misunderstanding due to differences in age, gender, cognitive ability, or interpretation of the question due to factors such as domestic or international citizenship.

Analysis also generated a graphic of the placement of persons and items on a common scale, which allowed for an examination of how the scale performed relative to the sample. The Rasch model graph for person-item fit simultaneously positions items and persons with respect to each other. This format was useful in viewing the extent to which items and persons match, and if the questions were appropriate for the persons. It also presents a visual summary of continuity by examining the gaps that suggest where items can be added or removed due to duplication, and if item order was appropriate.

Hierarchical Regression

Hierarchical regression (HR) is a form of multiple regression that involves a series of analyses. It can be thought of as building successive linear regression models, each adding more predictors. In this analysis, the same criterion was used and the independent variables were entered sequentially. This method allowed the researcher to determine which order to use for a list of predictors, which was achieved by putting the predictors into blocks of variables. The block can be a single predictor or a group of predictors. The first analysis contains one or more predictors, with the next analysis adding new predictors to those already used. The change in R^2 between the two analyses represented the proportion of variance in the criterion that was shared exclusively with the newly added variable(s). In the case of the data for this study, based on past research, differences between different groups, such as gender, age or ethnicity, were expected. Because of this, the use of HR was an appropriate method to use over standard regression. An important consideration in HR is the order of the variables. Since the effects of the variables entered in earlier steps are partialled from the relationships of the later steps, partial indices from different steps in the HR did not involve the same sets of variables, and were not directly comparable to one another.

Some of the assumptions of HR were that variables are approximately normally distributed, the relationship was linear in the parameters, variables were reliably measured, and homoscedasticity of error variance.

Latent Class Analysis

Latent class analysis (LCA) has seen increasing use in the social and health fields, and is another model-based approach. LCA is considered a subset of SEM, and is often referred to as finite mixture modeling (Vermunt & Magidson, 2002). It is used to cluster data into groups based on their responses to a set of observed variables. LCA is often used to uncover homogeneous groups based on observed variables; however, it relies on a hypothesized model. The latent structure may be either unidimensional or multidimensional. This flexibility allows researchers to test whether the measured variables define a unidimensional or multidimensional variable within the population, and can be used to test if the latent variable is invariant over multiple populations (McCutcheon, 1987). LCA is used for analyzing relationships among variables that are either nominal or ordinal.

LCA is similar to CFA, since both estimate latent variables from measured variables; however, in LCA, the latent variables are treated as categorical (groups). Continuous variables are termed factors, while categorical latent variables are termed latent class variables (Wang & Wang, 2012). Therefore, CFA groups items and is variable-centered. LCA groups respondents or cases based on patterns of responses, and is person-centered. Most LCA models are categorical; however newer software, such as *Mplus* allows for categorical, continuous, and count indicator variables (Wang & Wang, 2012). This is referred to as latent profile analysis (Muthen & Muthen, 2012). LCA can also be used for density estimation and for probabilistic cluster analysis for continuous, observed variables.

Similar to CFA, LCA uses latent variables to describe relationships between observed variables. LCA assumes local independence, which requires that the observed variables be mutually independent. LCA requires neither multivariate normality nor continuity of measurement. As with factor analysis, the model parameters are estimated for measurement errors; however it is considered the qualitative analog to factor analysis because it allows for the discovery of latent variables from two or more observed variables (McCutcheon, 1987). Two additional psychometric techniques based on LCA are latent trait analysis (LTA) and latent profile analysis (LPA). LTA allows for the characterization of continuous latent variables from discrete observed variables and, by extension, LPA allows for the characterization of discrete latent variables from continuous observed variables.

LCA estimation is based on maximum likelihood. To estimate a LCA model, several steps are followed: First, a latent class model is specified from a set of observed variables. Cases are then assigned to latent classes based on predictions. Finally, predicted scores are used to assess the class membership. Goodness of fit is usually tested by the Pearson or likelihood-ratio chi-squared statistic (Vermunt & Magidson, 2002). Different methods can be used to identify the best model fit: the likelihood ratio test, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), or the Lo-Mendell-Rubin adjusted likelihood ratio test. They measure the goodness of fit and the information lost with the different models when a given number of classes is used. After choosing the ideal number of classes, LCA then calculates probabilities for the presence of the latent variables for each class. Accuracy is measured by comparing the results of a new test method, with the probabilities calculated based on these classes of variables.

There are some restrictions and limitations for the unrestricted LCA model. Three or more categorical variables are needed. Only two latent classes can be identified with three dichotomous variables, and a greater number of dichotomous variables may lead to unidentified classes. Small sample size, correlated or continuous variables, and small (i.e., 0 or near 0) and large (i.e., equal to or near 1) probability estimates in distinct latent classes can lead to overestimation and underestimation of class prevalence, respectively (Neuhaus & Ring, 2013). A conditional item-response probability equal to 1.0 indicates that members in the latent class endorse an item.

Value of the Four Analysis Techniques

A strength of this study was the ability to leverage the complementary nature of EFA, IRT, regression, and latent class analysis to examine this new screening measure. While there were previous psychometric assessment of both the PHQ-4 and the AUDIT-C in general populations, there had been no assessment in a college-aged population. Further, there was a lack of assessment of impulsivity measures in this population, or in comparison to the PHQ-4 and AUDIT-C. This study blended the orientation of the statistical analyses by first examining the EFA's variable-oriented approach (Collins & Lanza, 2010), followed by IRT's item and person approaches (Linacre, 2010), then followed by a regression model approach, and concluding with latent class analysis's person oriented approach (Muthen & Muthen, 2012)

With its emphasis on identification and accounting for the linear relationships between observed variables across persons, EFA's variable oriented approach supported the anticipated multidimensionality of the complete scale, and the unidimensionality of the subscales (PHQ-4, AUDIT-C and the impulsivity questions). The Rasch analysis, with its capacity to estimate both the item difficulty and the person ability, provided support for the multidimensional structure of the full instrument, as well as supported the unidimensionality of the three subscales.

Hierarchical regression, another variable oriented approach, has the capacity to hold one level of predictors constant while adding levels based on theoretical relationships. Using this method, the researcher was able to determine the nature of the relationship and the contribution to the variance that was from the newly added variable.

The nature of LCA, with its focus on grouping unobserved subpopulations based on patterns of response, LCA's person oriented approach organized the latent structures of the data base on the probability of endorsing the different responses scales found in the instrument. With an emphasis on the individual, the results of the LCA provided a different perspective from EFA, Rasch, and regression regarding how to explain the differences in the item responses and how to characterize the response structure in a meaningful way.

Software

Analyses were performed using IBM® *SPSS*® statistics software (Version 22) for descriptive statistics, *Mplus* (Version 7.11) for LCA, and *Winsteps* (Version 3.81) for Rasch modeling.

Participants

The participants in the study were adult undergraduate and graduate students at a western university who visited a student Health and Counseling Center (HCC) during the Spring Quarter of 2014 (April to June). The Carnegie Foundation classifies this institution as a research university with high research activity. The student body is made up of approximately 5,100 undergraduate and 6,400 graduate students. The university population has slightly more female students than males (54% to 46%), and is approximately 8% international. The population that visits the HCC approximately mirrors the demographics of the greater university population. There are some small differences in the demographics of students that visit the HCC when compared to the university population. The HCC sees slightly more females than males, more graduate students than undergraduates, and slightly fewer international and ethnic minority students than the overall population (see Appendix E).

Consent for this intake instrument, as with other diagnostic medical and mental health instruments used in the HCC, was obtained at the first visit to the HCC each academic year by way of a comprehensive electronic consent and policy agreement by the student. The consent made the student aware that non-identifiable patient information gathered during visits to the HCC may be used for internal and external research, quality improvement, as needed by governmental or other reporting agencies, and for other purposes deemed necessary by the HCC. The student was also informed of his or her rights, responsibilities, and that, in cases where a risk was presented to the patient or to others, or as required by law, their identifiable information may be released on a limited basis.

During the collection period (Spring Quarter 2014), 491 individual questionnaires were captured and reviewed for analysis. Appendix D provides a description of the background characteristics of the 491 students in the sample.

Instruments

As stated earlier, while there are multiple diagnostic and screening instruments for depression, alcohol use, and impulsivity, there is currently no brief screening tool that captures depression and alcohol misuse that is specifically related problem drinking in one instrument. There is also no brief screening instrument that explores possible impulsivity components related to depression and problem drinking that would assist with early identification and intervention. The purpose of this research study was to investigate the effectiveness of a newly developed, 15-question intake screening tool for use in a primary care medical setting. The new screening tool combined items from two established and validated measures used to screen for depression: the PHQ-4 and the AUDIT-C.

Both the PHQ-4 and the AUDIT-C have been validated in college-aged and in general populations; however, one of the challenges with alcohol screening in a college population is the number of students that score high on the alcohol scale. It was anticipated that even with a higher cut-off score on the AUDIT-C questions, over 50% of students answering the questions would need some sort of alcohol use intervention. Another issue was that the resulting score is based on consumption, not on any underlying AUD. This made decisions about what intervention the medical provider should use difficult without increased screenings. Having a health provider handle this volume of high scores on the AUDIT-C is not possible in a busy medical practice.

The challenge was to include questions or develop a scale that could differentiate between high-risk drinking (such as excess consumption and/or binge drinking) from excessive drinking that could be diagnosed as an AUD or that could (or has) become a life-long problem. Since most college-aged excess consumption patterns are mature during or just after the college experience, it was anticipated that using additional questions that screen for impulsivity facets that correlate with these alcohol issues would aid in identifying students the most at risk for life-long alcohol-related issues.

The screening tool was also intended to investigate if a targeted number of impulsivity questions related to depression and anxiety, and could help with early identification of suicidal ideation or advanced depression; however, the research in this area is less explicit.

Developing a Brief Comprehensive Screening Instrument

The 15-question screening instrument was comprised of the PHQ-4 (which consists of four questions related to anxiety and depression), the three questions of the AUDIT-C, and eight impulsivity questions from the UPPS-P impulsivity scale. Due to the medical practice software being used at the HCC, there was a limit placed on the number of questions that could be presented on an intake questionnaire. While this limitation was difficult, another consideration was to limit the number of questions at intake to an amount that could be answered in a short period of time before the visit. The currently used scale, the PHQ-9, takes less than two minutes to complete. The desire of the clinic was to keep the time needed to complete the new instrument to less than four minutes, which is why the limit of approximately 15 questions was chosen.

The PHQ-4 questions were kept in their entirety, since this scale has been well validated, is psychometrically sound, and the equivalent, PHQ-9, is currently in use as part of the medical patient intake process. The AUDIT-C was chosen because it has similar psychometric characteristics as the longer 10-question AUDIT, and has been found to be effective in screening for alcohol frequency issues—a key component of problem and binge drinking (Aalto, Alho, Halme, & Seppä, 2009; Meneses-Gaya et al., 2010).

The eight experimental questions on impulsivity were chosen based on high factor loadings on the impulsivity factors that were most related to the factors being investigated by the PHQ-4 and AUDIT questions. Research on these items has been previously detailed, and is based on work by Billieux et al. (2012) from their research on the short French version of the UPPS-P instrument. In this research article, factor loadings of the 20 items were detailed, and for each impulsivity facet identified, the strongest loading items were selected. Due to the limitation on the total number of questions, when two items were loaded on the same impulsivity facet, only one item for each facet was identified.

Initially, only the PHQ-4 and the AUDIT-C were to be scored and evaluated by the healthcare provider for an intervention. Using the cut-offs for intervention recommended by previous studies, a score of five or above was used on the PHQ-4 (Löwe, 2010). The recommended cut-off for intervention on the AUDIT-C in a normal population is a score greater than four for men, and three for women (Aalto et al., 2009; Dawson, 2012). Because of the prevalence of excessive consumption of alcohol on a college campus, a previously reported cut-off of five for men and four for women was used initially (Bradley et al., 2007; Graham, 2007).

Procedure

Approval to conduct the study was granted from the University of Denver Institutional Review Board (IRB). Permission to administer the measure was granted by the University of Denver Health and Counseling Center. Participants were selected from primary care medical visits to the HCC at a western university, and were required to take the new instrument that consisted of the PHQ-4, the AUDIT-C, and eight questions from the UPPS-P (see Appendix A). The questionnaire was presented electronically at a computer kiosk upon check-in. dependent on the type of visit scheduled. These visits were typically same-day appointments for general illnesses and well visits. Certain visits were excluded, such as a visit within the prior week, urgent care visits, allergy and immunization injection, and mental health-related visits. Unless the student was extremely late for their scheduled visit, or was excluded by type of visit, the student was required to complete the questionnaire. The questionnaire was only presented with certain medical providers that were trained in interpreting the results of the questionnaire.

Scores were interpreted only at the time of the visit for the PHQ-4 and the AUDIT-C, with the remaining eight questions used for instrument development. Questionnaire results were reviewed for total score initially by a medical assistant when the student began the appointment, and then by the medical provider. Cutoff scores for provider intervention were four for the PHQ-4 and six on the AUDIT-C. When appropriate, scores on the other eight questions were referred to for any additional information. The HCC also has a behavioral health consultant available, who is a psychologist to assist with students with scores above a set level.

CHAPTER THREE: RESULTS

This chapter reports the results of the assessment of the PHQ-4, the AUDIT-C, and impulsivity items through the application of exploratory factor analysis (EFA), item response theory (IRT), hierarchical regression (HR), and latent class analysis (LCA). Results are based on the questions posed in Chapter 1.

Research Questions

Research Question One

Exploratory factor analysis was used to determine if the new scale has a unidimensional structure. It was hypothesized that the structure is multidimensional.

The item level responses were examined for underlying patterns via factor analysis procedures using *SPSS*. The data were initially screened for normality, univariate outliers, and missing data. A prerequisite for including an item was that responses were not too badly skewed (i.e., 90% or more of responses clustered in single cell) and, more generally, that the level of response to that item was sufficient (<15%-20% missing) to destabilize analysis. After examination, all items were included in the initial factor analysis.

Initially, the protocol used for the exploratory factor analysis was principal axis factoring (PAF) and a rotation of the matrix of loadings to obtain orthogonal (independent) factors (i.e., Varimax rotation). Since there were theoretical grounds on which the factors might correlate with each other, rotation was repeated using an oblique rotation. Additionally, principal components analysis (PCA) was used with orthogonal and oblique rotation methods, with the factor structure being compared to results from a PAF. In most instances (including this one), PCA and PAF yield similar results, but because the PAF method focuses on shared variance and not on sources of error, it has been deemed more appropriate for use in the social and behavioral sciences. The prime goal of factor analysis is to identity a simple structure (items loadings >0.30 on only one factor) that is interpretable, assuming that items are factorable. The Kaiser-Meyer-Olkin measure of sampling adequacy was used to determine if the partial correlations among variables were adequate for factoring. Bartlett's test of sphericity was examined to determine if the correlation matrix was or was not appropriate. Factor loadings greater than 0.10 were examined, even though only item loadings over 0.30 were considered relevant for interpretation (i.e., as an item that reflected a factor). Several models with different rotations were examined to determine the underlying factor structure that was most interpretable. Overall, all generated models displayed a similar factor structure, and when rotated with different methods again, resulted in similar factor structures.

Initially, a PAF with a Varimax (orthogonal) rotation of the 15 questions from the measure was conducted on data gathered from 491 participants. Several well-recognized criteria for the factorability of a correlation matrix were used. First, all 15 items correlated at least .3 with at least one another item, suggesting reasonable factorability. Second, an examination of the Kaiser-Meyer Olkin measure of sampling adequacy suggested that the matrix was factorable and also that the recommended minimum value of .6 was exceeded (KMO ranged from .75 to.77 on the different models). Bartlett’s test was used to test the null hypothesis that the original correlation matrix was an identity matrix. In all of the models examined, the Bartlett test result was statistically significant ($p < 0.001$); therefore, factor analysis was determined to be appropriate. Finally, 12 of the 15 items had communalities that were above 0.3, confirming that the items shared some common variance with other items.

The initial factor analysis had eigenvalues indicating that the first factor explained 26.47% of the variance, the second factor explained 16.80% of the variance, the third factor explained 14.04% of the variance, and the fourth factor explained 8.8% of the variance. The fifth and sixth factors had eigenvalues of just below one, and each factor explained 6% and 5% of the variance, respectively. Parallel analysis supported a three-factor solution; however, three- and four-factor solutions were examined, using both PC and PAF with Varimax and oblimin rotations of the factor loading matrix. The three-factor solution, which explained 57.32% of the variance, was preferred because of support from parallel analysis, examination of the scree plot, and from theoretical support, with factors reflecting the two validated scales (the PHQ-4 and the AUDIT-C) and the impulsivity questions from the UPPS-P. The fourth and subsequent factors were not interpretable. There was little difference between the orthogonal and oblique solutions; thus, both solutions were examined in the subsequent analyses before deciding on an oblique rotation for the final solution.

As anticipated, the final factor solution was multidimensional. The final factor solution used PAF with oblique rotation of 14 of the 15 Likert-scale questions from the screening questionnaire. One question, number 27 of the UPPS-P scale, was dropped since it did not have a loading over .30 on the three retained factors. The final result constrained the number of factors to three, and used oblique rotation to enhance interpretability. Loadings under .3 were not displayed.

As shown in Table 1 and Table 2, in both the pattern and structure matrix, four items loaded on factor one, and three items loaded on factor two. The four items are the questions that made up the PHQ-4, and the three items that made up the second factor are the AUDIT-C questions. Four of the remaining six questions from the UPPS-P loaded on the third factor in both the pattern and structure matrix. The remaining two UPPS-P items had values below .3 in the pattern matrix, but were present in the structure matrix in either factor one or factor three. The factor correlation matrix indicated that the three factors were not correlated (range -0.26- 0.32). Since this instrument was made up of two validated instruments, the PHQ-4 and the AUDIT-C, as well as questions from the UPPS-P, this result was somewhat expected. In reflecting on the makeup of this questionnaire, the three factors identify level of depression, alcohol use, and impulsivity. This structure corresponds to the underlying theory that was used to develop this pilot screening instrument. This factor structure was also useful when performing the Rasch analysis.

Table 1

Exploratory Factor Analysis Factor Loadings for 15-item Measure (N = 491)

	Factor		
	1	**2**	**3**
PHQ-4 Q1	.58	-.38	
PHQ-4 Q2	.67	-.43	
PHQ-4 Q3	.57	-.33	
PHQ-4 Q4	.64	-.34	
AUDIT-C Q1		.43	.44
AUDIT-C Q2		.55	.37

Table 1 (continued)

	Factor		
	1	**2**	**3**
AUDIT-C Q3		.72	.53
UPPS-P Q48	.40		
UPPS-P Q50	.39	.34	
UPPS-P Q53	.68		-.44
UPPS-P Q29	-.31		
UPPS-P Q44	.64		-.45
UPPS-P Q41	.60		-.48
UPPS-P Q31			
Eigenvalue	3.71	2.35	1.97
% of Variance	26.47	16.80	14.04

Note: Factor loadings under .30 are not shown.

Table 2

Exploratory Factor Analysis Pattern Matrix for 15-item Measure (N = 491)

Pattern Matrix			
	Factor		
	1	**2**	**3**
PHQ-4 Q1	.75		
PHQ-4 Q2	.85		
PHQ-4 Q3	.73		
PHQ-4 Q4	.78		
AUDIT-C Q1		.63	
AUDIT-C Q2		.68	
AUDIT-C Q3		.91	
UPPS-P Q48			
UPPS-P Q50			-.47
UPPS-P Q53			-.81

Table 2 (continued)

Pattern Matrix			
	Factor		
	1	2	3
UPPS-P Q29			
UPPS-P Q44			-.81
UPPS-P Q41			-.83
UPPS-P Q31			
Eigenvalue	3.27	1.91	1.67
% of variance	23.34	13.66	11.22

Note: Extraction Method: Principal Axis Factoring; Rotation Method: Oblimin with Kaiser Normalization; Factor loadings under .30 are not shown.

Research Question Two

Rasch analyses of the full scale and potential subscales were examined to determine if there was a dimensionally appropriate structure, appropriate scale use, and presence of differential item functioning by gender, year in college, and ethnicity.

Analyses were performed using IBM *SPSS* statistics software (Version 22) for descriptive statistics, and *Winsteps* (Version 3.81.0; Linacre, 1999-2014) for Rasch analyses. Since the survey responses were entered into a computer terminal, there were no missing or invalid data for items or persons. Descriptive analysis was performed to establish that there were no outliers and that the data met the assumptions for Rasch analysis. The full instrument was developed to screen for a multitude of aspects of the latent constructs of depression, alcohol abuse, and impulsivity. With all 15 items examined by the EFA, it was expected that multidimensionality would be present.

Research questions number three, four, and five (respectively) address dimensionality of the depression, alcohol use, and impulsivity factors identified in the EFA.

Dimensionality

Dimensionality was examined in this study using: (a) overall fit of the data to a one-dimensional model; (b) a Rasch principal components analysis of residuals; and (c) individual item fit.

Fit was measured in at least two ways in the Rasch model: infit (a measure weighted by the distance between person and item location) and outfit (the un-weighted measure). It was expected that the mean square (MNSQ) indices were 1.0, if the data fit the model exactly. Acceptable infit and outfit values for items with a Likert-scale response commonly fall within a range of 0.6 to 1.4 (Walker et al., 2012). For the full scale, the average MNSQ for person infit was 1.00 and 1.04 for outfit, which indicates adequate fit (see Table 3). The average MNSQ for item infit was 1.08 and 1.04 for outfit, which also indicates adequate fit, but with the potential for misfitting items or persons (see Table 3). However, when examining individual item fit, only seven of the fifteen items displayed proper fit—with four items displaying overfit, and four items underfit. The item with the worst fit was UPPS-P question 27. This suggests that, as a set, the 15 items did *not* reflect a unidimensional construct.

Dimensionality was also examined using a Rasch principal components analysis of residuals. The results (shown in Table 3) indicated multidimensionality, with approximately 68.7% of the variance explained by the measure. The first contrast eigenvalue was 3.2. It is recommended that the eigenvalue for the first contrast should be less than 2.0 to be considered unidimensional (Linacre, 2010). The data were then examined after removing UPPS-P question 27, which displayed the poorest infit/outfit. The overall results improved slightly, but continued to show multidimensionality based on a first contrast eigenvalue that exceeded 2.0.

Table 3

Dimensionality, Fit, and Separation for the 15-Item Measure (PHQ-4, AUDIT-C, and 8 Impulsivity Items)

Number of items	15
Dimensionality—eigenvalue for 1st contrast	3.2
Mean Item MNSQ Infit	1.08
Mean Item MNSQ Outfit	1.04
Mean Person MNSQ Infit	1.00
Mean Person MNSQ Infit	1.04
Real Person Separation	1.10
Real Reliability of Person Separation	.55
Cronbach's Alpha	.60
Person Logit Mean	-.78

Finally, dimensionality using individual item fit was examined to see if any items misfit the Rasch model. Mean square infit and outfit were examined using methods previously described. Infit mean square ranged from 1.94 to .38. Outfit mean square ranged from 2.16 to .37. Based on both mean square infit and mean square outfit, the items were misfitting. It was concluded that when using all items, the scale was multidimensional. Therefore, dimensions identified in the EFA were examined individually.

Research Question Three

Rasch analysis of the PHQ-4 was used to examine the dimensional structure in a college-aged population and the presence of differential item functioning by gender, year in college, and ethnicity.

It was expected that the four items identified by EFA that comprised the anxiety and depression factor would reflect a unidimensional construct. These four items are also the four items of the PHQ-4 instrument.

Dimensionality

For the PHQ-4 scale, the average MNSQ for person infit was .97 and 1.00 for outfit, which indicates adequate fit. The average MNSQ for item infit was 1.04 and 1.00 for outfit, which also indicates adequate fit (see Table 4).

Dimensionality was examined using a Rasch principal components analysis of residuals for the calibration and validation samples. The results indicated likely unidimensionality, with approximately 58.1% of the variance explained by the measure. The first contrast eigenvalue was 2.1, which is slightly above the value of 2.0 to be considered unidimensional (Linacre, 2010). The PHQ-4 is constructed to measure depression and anxiety: the first two items on the instrument measure anxiety, and the final two questions measure depression. While the comorbidity of these two mental illnesses is usually over 50%, the fact that these two issues are not always co-occurring might be the reason for the slight elevation of the first contrast eigenvalue.

Table 4

Dimensionality, Fit, and Separation for the PHQ-4

Number of items	4
Dimensionality—eigenvalue for 1st contrast	2.1
Mean Item MNSQ Infit	1.04
Mean Item MNSQ Outfit	1.00
Mean Person MNSQ Infit	.97
Mean Person MNSQ Outfit	1.00
Real Person Separation	.97
Real Reliability of Person Separation	.48
Cronbach's Alpha	.85
Person Logit Mean	-4.21

Finally, dimensionality using individual item fit was examined to see if any items misfit the Rasch model. Mean square infit and outfit were examined using methods previously described. Individual item fit was also examined, and all items had acceptable infit and outfit: between .7 and 1.4. It can be concluded that when using the four items, the scale was substantially unidimensional.

Reliability

Both person and item separation and reliability of separation measure instrument spread across the trait continuum. Reliability of person separation is conceptually similar to Cronbach's alpha, though it is generally lower because it is computed without extreme scores. Extreme scores are removed when doing this analysis, since they cannot be accurately located on the trait. Reliability of separation measures the spread of items and persons in standard error units. To be useful, instruments should have a separation of at least 2.0. Higher values indicate a wider spread of items and persons, and lower values indicate less separation. Separation determines reliability, with higher separation of persons or items yielding higher reliability. Person separation for this study was .97, with a Cronbach's alpha of .85, indicating that the PHQ-4 had some separation and was a marginally useful instrument for diagnosing depression with this sample. When removing extreme persons, person separation was 1.24, with an alpha of .61 (229 non-extreme persons). The relatively low person reliability was expected with the PHQ-4, since the PHQ-4 is a depression and anxiety screening tool, and most of the persons being measured were not depressed.

Invariance

Invariance is critical to the usefulness of the PHQ-4 in screening college-aged individuals for depression. Invariance is defined as consistency in the ordering of item responses across person groups with differing characteristics. Examples of invariant measures include height, weight, and temperature. Measures of health, such as pain or depression, tend to be fairly invariant. Failure of invariance means that it is inappropriate to compare samples because the construct is construed differently by each sample. Bond and Fox (2007) defined invariance as the maintenance of identity in the meaning of a variable from one time or group to the next.

Invariance can be assessed with differential item functioning (DIF), which is based on a shift in meaning that is expressed as item location over time or between groups. DIF may occur due to construct misunderstanding due to differences in age, gender, or cognitive ability. DIF may also occur due to interpretation of the question due to factors such as domestic or international citizenship.

Invariance of the PHQ-4 was assessed by computing DIF by gender, age, and ethnicity. As noted in Chapter 2, significance of DIF was assessed by generating the difference in logit position by group or time, and dividing the difference by the combined standard error. As shown in Table 5, for gender, there was no DIF on items 1, 2, and 4; there was DIF for item 3, which is a depression question. Item 3 was easier for females to agree with than it was for males (see Table 5; logit position for females = 1.10; logit position for males = 0.30).

Table 5

Differential Item Functioning by PHQ-4 Logit Item Position

	Mean Logit Person Position		
Gender	Female	Male	*p*
Item #3	1.10	.30	.08*

Note: $N = 491$. $*p < .05$ $**p < .01$ $***p < .001$

DIF for age was determined by comparing individuals aged 22 or younger (i.e., undergraduate students) with individuals aged 23 and older (i.e., graduate students); there was no DIF between the age groups. For ethnicity, there was no DIF between white and non-white individuals.

According to Linacre (2010), tests of significance (including DIF tests), when done in a Rasch context, are of uncertain value, as differences can be statistically significant but too small to impact meaning or the practical use of the measures. As such, statistical significance and substantive difference are required to take action against bias. In this case, there was a substantial difference between the logit positions; it was therefore concluded that DIF was present by gender, and should be investigated further. However, for the current study, the PHQ-4 logit person position was used across genders.

Scale Use and Targeting

Scale use was as expected for the PHQ-4 with no inversions in the step structure, as depicted in Table 6 and in Figure 1. In Figure 1, the curves show how probable each category was to observe relative to the item measured, which is expressed as the difference between item and person logit position. The probability of response is the likelihood of endorsing a given rating-scale category at that level of difference in the person item of depression. The intersection of adjacent rating scale categories can be seen at an estimated threshold value of the higher of the two categories.

The placement of persons and items on a common scale allowed for an examination of how the scale performs relative to the sample. The *Winsteps* software graphs item location with person location. The Rasch model graph for person item fit simultaneously positions items and persons with respect to each other. This format is useful in viewing the extent to which items and persons match, and if the questions are appropriate for the persons. It also presents a visual summary of continuity by examining gaps that suggest where items can be added or removed due to duplication, and if item order is appropriate. When examining the PHQ-4 for these items, it was not expected that items and persons would have similar means, and it was expected that the majority of participants would have very low scores. According to previously cited research, approximately 10% of this population is diagnosed with depression; therefore, individuals with less traits of depression were expected to score lower using this screening tool. Additionally, some items in the scale may have been easier to endorse than others. For example, question 1, “Feeling nervous, anxious, or on edge” was more likely to receive a higher score than question 4, “Feeling down, depressed or hopeless.” By reviewing Figure 2, how well the measure targeted the persons who participated in the study can be seen. The left side of the figure represents persons. Each “#” represents 21 individuals, and “.” represents 20 individuals. Based on this distribution, it appears that a small percentage of this population was measured well by the instrument, as the majority of the sample fell below -5.0 on the scale. This suggests the general absence of anxiety and depression in this sample population. This placement was expected, since this is an anxiety and depression screening tool and the population was expected to have an absence of depression. Item order was as anticipated.

Table 6

Step Structure for PHQ-4

Category	Observed Percentage	Observed Average	Infit MNSQ	Step Structure
0	72	-3.88	1.01	NONE
1	22	-1.77	1.03	-3.01
2	4	.42	.95	1.00
3	2	2.1	1.20	2.01

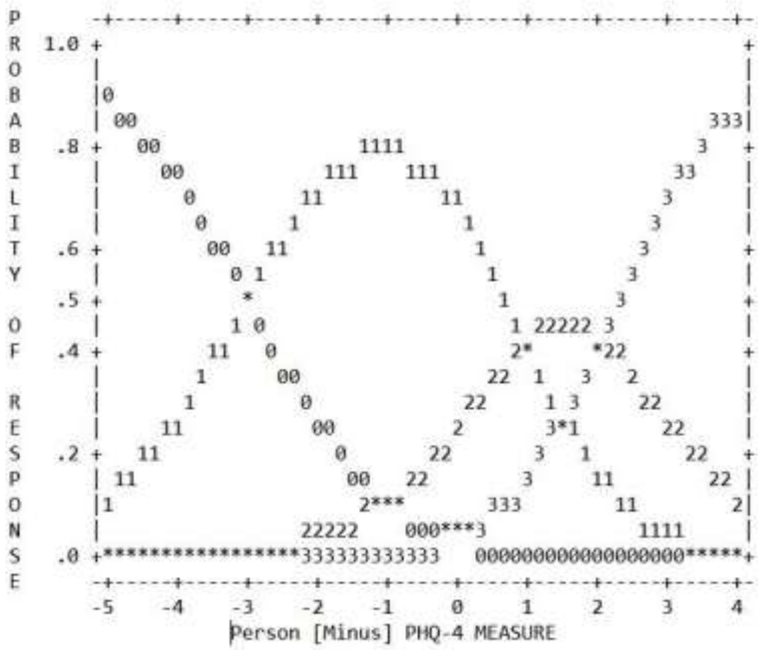

Figure 1. PHQ-4 rating scale use.

```
---------------------------------------------------------------------------------
MEASURE in LOGITS      | BOTTOM P=50%   | MEASURE       | TOP P=50%    MEASURE
  <more> ----- PERSON -+- ITEM         -+- ITEM        -+- ITEM          <rare>
   4                 . +                +               +                   4
                       |                |               |
                     . |                |               |
                       |                |               | PHQ4-4
                       |                |               |
                       |                |               |
   3                   +                +               + PHQ4-3            3
                     . |                |               |
                       |                |               |
                       |                |               |
                       |                |               |
                     . |                |               |
   2                   +                +               +                   2
                       |                |               | PHQ4-2
                       |                |               |
                     . |                |               |
                       |                |               |
                       |                | PHQ4-4        |
   1                 . +                +               +                   1
                       |                |               | PHQ4-1
                       |                | PHQ4-3        |
                       |                |               |
                     . |                |               |
                       |                |               |
   0                   +                +               +                   0
                       |                |               |
                     . |                |               |
                       |                | PHQ4-2        |
                       |                |               |
                       |                |               |
  -1                   +                +               +                  -1
                    .# |                |               |
                       |                |               |
                       |                | PHQ4-1        |
                       |                |               |
                       | PHQ4-4         |               |
  -2               .## +                +               +                  -2
                       |                |               |
                       | PHQ4-3         |               |
                       |                |               |
                       |                |               |
                       |                |               |
  -3                   +                +               +                  -3
                   .## |                |               |
                       |                |               |
                       | PHQ4-2         |               |
                       |                |               |
                       |                |               |
  -4                   +                +               +                  -4
                       |                |               |
                  .### |                |               |
                       | PHQ4-1         |               |
                       |                |               |
                       |                |               |
  -5     .############ +                +               +                  -5
  <less> ----- PERSON -+- ITEM         -+- ITEM        -+- ITEM          <frequent>
EACH "#" IN THE PERSON COLUMN IS 21 PERSON: EACH "." IS 1 TO 20
```

Figure 2. Map of person and item for PHQ-4.

Although DIF was present on one question for gender (question 3), this set of items was treated as a measure.

Research Question Four

Rasch analysis of the impulsivity factor identified by EFA was used to examine the dimensional structure in a college-aged population.

With all eight items examined by EFA, it was expected that unidimensionality would be present.

Dimensionality

For the impulsivity factor, the average MNSQ for person infit was .98 and 1.04 for outfit, which indicates good fit. The average MNSQ for item infit was .97 and 1.04 for outfit, which also indicates good fit.

Dimensionality was also examined using a Rasch principal components analysis of residuals for the calibration and validation samples. The results indicated possible multidimensionality, with approximately 57.8% of the variance explained by the measure. The first contrast eigenvalue was 2.9, which is above the value of 2.0 to be considered unidimensional (Linacre, 2010).

Finally, dimensionality using individual item fit was examined to see if any items misfit the Rasch model. Mean square infit and outfit were examined using methods previously described. Individual item fit was examined, and four items had unacceptable infit and outfit. Since the items selected for this scale were from a much larger scale that measured five aspects of impulsivity related to either suicidal ideation or a variety of alcohol misuse issues, this was expected. Items were removed one at a time by removing the under-fitting items and leaving the over-fitting items determine if a core of the impulsivity items formed a unidimensional scale with better fit.

After removing four poorly fitting items (items 1, 2, 5, and 8), a four-item scale remained that met the criteria for unidimensionality. The remaining UPPS-P questions were 41, 44, 50, and 53. The average MNSQ for person infit was .98 and .98 for outfit, which indicates good fit (see Table 7). The average MNSQ for item infit was .98 and .98 for outfit, which also indicates good fit. The four remaining items were theoretically appropriate based on prior associations between impulsivity and drinking behavior.

The Rasch dimension explained 58.8% of the variance in the data. Results with the modified scale indicated possible unidimensionality, with approximately 18.2% of the variance explained by the first contrast. The first contrast eigenvalue was 1.8, which is below the value of 2.0 to be considered unidimensional (Linacre, 2010).

Table 7

Dimensionality, Fit, and Separation for Impulsivity Measure

Number of items	4
Dimensionality—eigenvalue for 1st contrast	1.8
Mean Item MNSQ Infit	.98
Mean Item MNSQ Outfit	.98
Mean Person MNSQ Infit	.98
Mean Person MNSQ Outfit	.98
Real Person Separation	1.67
Real Reliability of Person Separation	.74
Cronbach's Alpha	.80
Person Logit Mean	-2.05

Reliability

Person separation was 1.67, with a Cronbach's alpha of .80, which indicates that the impulsivity factor had some separation and that the instrument was useful for diagnosing impulsivity, as impulsivity was most likely associated with alcohol consumption within this sample. When removing extreme persons, person separation was 1.34, with an alpha of .64 (414 non-extreme persons).

Invariance

Invariance is critical to the usefulness of the impulsivity factor in screening college-aged individuals for impulsivity. Invariance of the impulsivity factor was assessed by computing DIF by gender, age, and ethnicity. There was DIF on one item for gender and for two items for age. DIF for gender was examined by comparing males and females. There was DIF on one item: UPPS-P question 50, which asks, "When I am really excited, I tend not to think about the consequences of my actions" (see Table 8; logit position for females = 1.57; logit position for males = 1.93, which yielded a relatively small DIF contrast).

DIF for age was determined by comparing individuals 22 years or younger (i.e., undergraduate students) with individuals 23 years and older (i.e., graduate students). Table 8 shows that there was DIF on two items: UPPS-P question 44, which asks if acting impulsively often makes matters worse (logit position for age group 0 = 1.84; logit position for age group 1 = 1.39); and UPPS-P question 50, which asks "When I am really excited, I tend not to think on the consequences of my actions" (logit position for group 0 = -1.00; logit position for group 1 = -.44). In both cases, DIF was statistically significant. For ethnicity, there was no DIF. The four-item impulsivity set was treated as a measure for all cases with significant DIF, but with relatively small DIF contrasts.

Table 8

Differential Item Functioning by Impulsivity Logit Item Position

Gender	Female	Male	*p*
Item #1 (UPPS-P #50)	1.57	1.93	.03*
Age	22 and younger	23 and older	*p*
Item #1 (UPPS-P #44)	1.84	1.39	.03**
Item #4 (UPPS-P #50)	-1.00	-.44	.00***

Note: $N = 491$. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Scale Use and Targeting

Scale use was as expected, with no inversions in the step structure for items (see Table 9 and Figure 3).

When examining the impulsivity factor for these items, it was expected that the items and persons would line up, since the selected (and remaining) questions in the three-item instrument were based on prior research on impulsivity and alcohol consumption. Figure 4 shows how well the measure fits the persons, and items can be determined. The left side of the figure represents persons in the calibration group. Each "#" represents seven individuals and each "." represents six individuals. Based on this distribution, it appears that the majority of this population was measured well by these questions, and the minority of the sample fell below -5.0 on the scale. This suggests that impulsivity is measured well in this sample. This placement was expected, since research has shown that younger college students tend to be very impulsive. Item loading order functioned as anticipated.

Table 9

Step Structure for Impulsivity Measure

Category	Observed Percentage	Observed Average	Infit MNSQ	Step Structure
1	40	-3.41	.99	NONE
2	31	-1.68	192	-2.76
3	24	.15	.95	.06
4	4	2.10	1.28	3.37

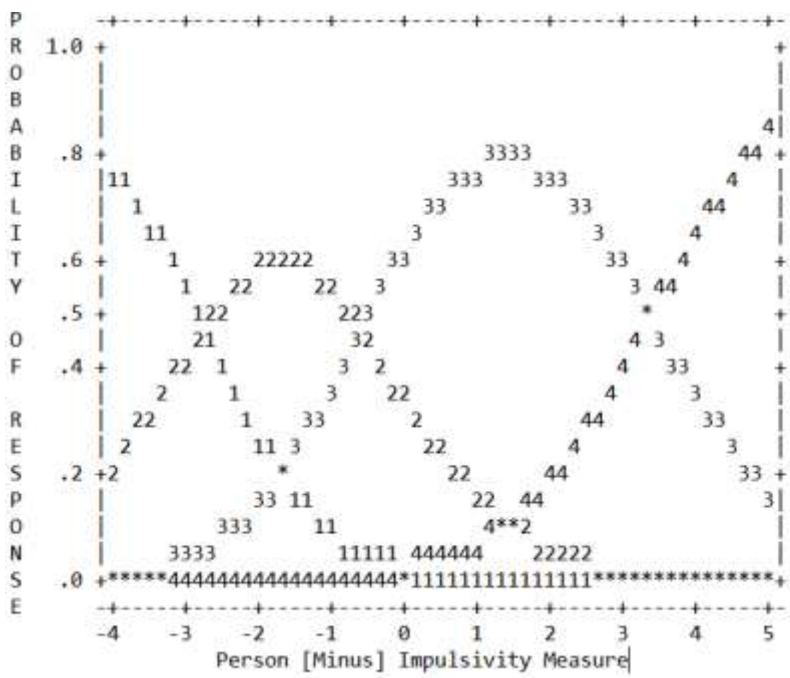

Figure 3. Impulsivity rating scale use.

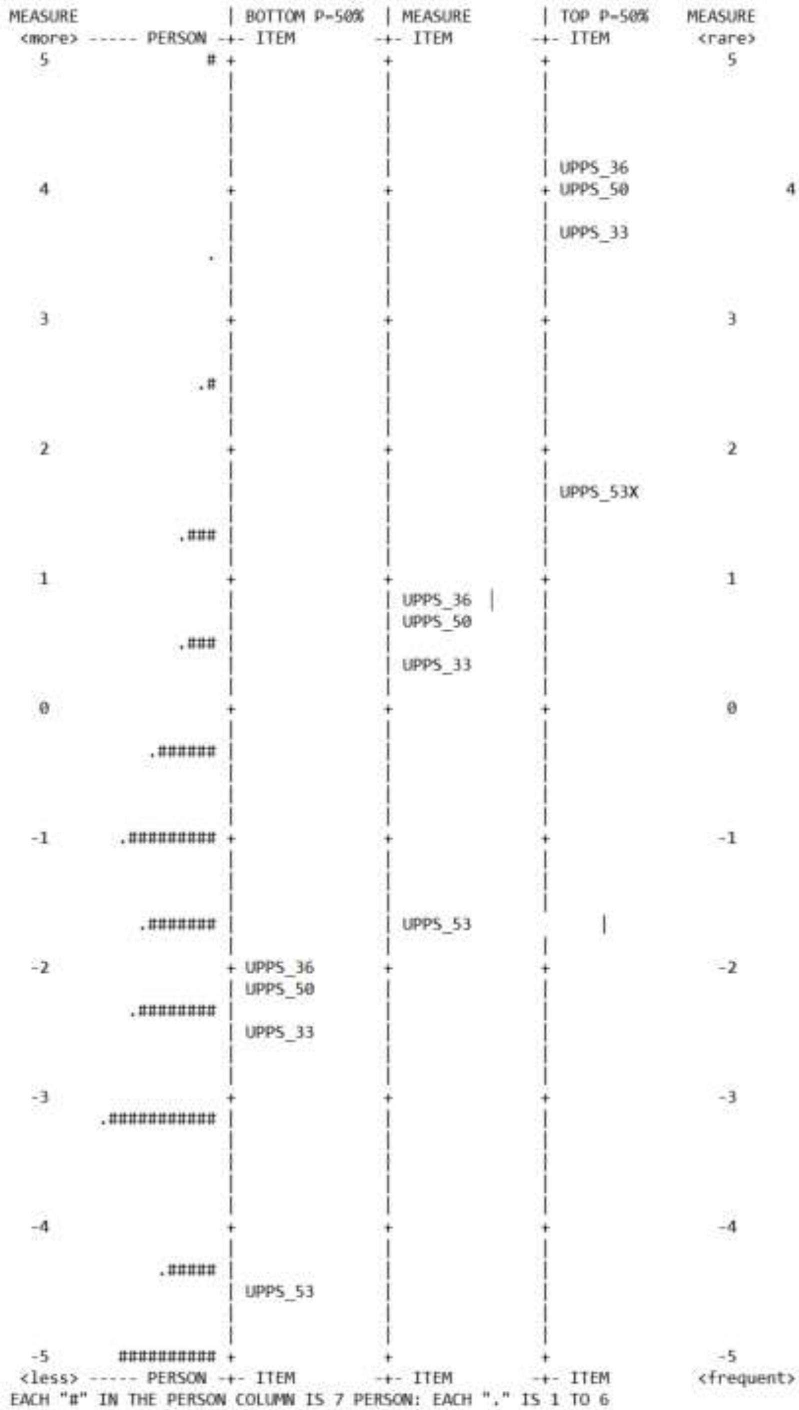

Figure 4. Map of person and item for four-item Impulsivity measure.

Although DIF was present on one question for age and two for gender, and there was no DIF for ethnicity, DIF contrasts were relatively small; therefore, this set of items was treated as a measure.

Research Question Five

Rasch analysis of the AUDIT-C was used to examine the dimensional structure in a college-aged population.

It was expected that unidimensionality would be present with all three items indicated as the second factor by EFA. These three items make up the AUDIT-C instrument.

Dimensionality

For the AUDIT-C scale, the average MNSQ for person infit was .92 and .97 for outfit, which indicates good fit. The average MNSQ for item infit was .95 and 1.01 for outfit, which also indicates good fit (see Table 10).

Dimensionality was examined by using a Rasch principal components analysis of residuals for the calibration and validation samples. The results indicated possible multidimensionality, with approximately 68.6% of the variance explained by the measure. The first contrast eigenvalue was 2.4, which is above the value of 2.0 needed to be considered unidimensional (Linacre, 2010).

Finally, dimensionality using individual item fit was examined to see if any items misfit the Rasch model. Mean square infit and outfit were examined using methods previously described. Individual item fit was also examined, and all items had acceptable infit and outfit: between .75 and 1.13. It was concluded that when using the three items, the scale could be treated as unidimensional. Further analysis of reliability, invariance, targeting, and scale use should be conducted.

Reliability

Person separation was 1.84 with a Cronbach's alpha of .77, indicating that the AUDIT-C had some separation. When removing extreme persons, person separation was 1.55 with an alpha of .71 (441 non-extreme persons).

Table 10

Dimensionality, Fit, and Separation for the AUDIT-C Measure

Number of items	4
Dimensionality—eigenvalue for 1st contrast	2.4
Mean Item MNSQ Infit	.95
Mean Item MNSQ Outfit	.97
Mean Person MNSQ Infit	.92
Mean Person MNSQ Outfit	.97
Real Person Separation	1.84
Real Reliability of Person Separation	1.00
Cronbach's Alpha	.77
Person Logit Mean	-3.40

Invariance

Invariance is critical to the usefulness of the AUDIT-C in screening college-aged individuals for alcohol consumption. Invariance of the AUDIT-C was assessed by computing DIF by gender, age, and ethnicity. For gender, there was DIF on item 1 (see Table 11; logit position for females = -3.17; logit position for males = -2.35) and item 3 (logit position for females = 1.15; and logit position for males = .31), with no DIF for item 2. This indicates that items changed difficulty position for males and females.

DIF for age was determined by comparing individuals 22 years or younger (i.e., undergraduate students) with individuals 23 years and older (i.e., graduate students). There was DIF on item 1 (logit position for females = -2.66; logit position for males = -3.33) and item 2 (logit position for females = 1.82; logit position for males = 2.94). For ethnicity, there was no DIF between white and non-white individuals.

Table 11

Differential Item Functioning by AUDIT-C Logit Item Positions

Gender	**Female**	**Male**	***p***
Item #1	-3.17	-2.35	.0001***
Item #3	1.15	.31	.0001***
Age	22 and younger	23 and older	*p*
Item #1	-2.66	-3.33	.0005***
Item #2	1.82	2.94	.0001***

Note: $N = 491$. *$p < .05$. **$p < .01$. ***$p < .001$.

Because DIF was both statistically significant and substantial, it was concluded that the AUDIT-C could not be reasonably treated as an invariant measure. The AUDIT-C was not treated as a scale; therefore, further analyses individually used items in the measure.

Scale Use and Targeting

Item order functioned as anticipated, and scale use was as expected for the AUDIT-C, with no inversions in the step structure (see Table 12 and Figure 5).

The placement of persons and items on a common scale allowed for an examination of how the scale performed relative to the sample. When examining the AUDIT-C for these items, it was expected that items would be adequately targeted, since the AUDIT-C is a diagnostic screening tool for alcohol consumption. It was also expected that the majority of participants would have a high consumption pattern. According to previously cited research, approximately 50% of this population drink (and also binge drink) regularly. Additionally, some items on the scale are simpler to endorse than others. For example, question 1, "How often do you have a drink containing alcohol?" is more likely to receive a higher score than is question 3, "How often do you have six or more drinks on one occasion?" By reviewing Figure 6, how well the measure fits the persons and items can be determined. The left side of the figure represents persons in the calibration group. Each "#" represents seven individuals, and each "." represents six individuals. Based on this distribution, it appears that a large percentage of this population was measured reasonably well by the instrument.

Table 12

Step Structure for AUDIT-C

Category	Observed Percentage	Observed Average	Infit MNSQ	Step Structure
0	40	-5.66	1.22	NONE
1	27	-3.38	.83	-4.51
2	17	-.60	.82	-1.54
3	13	1.81	.83	.69
4	2	4.36	1.75	5.00

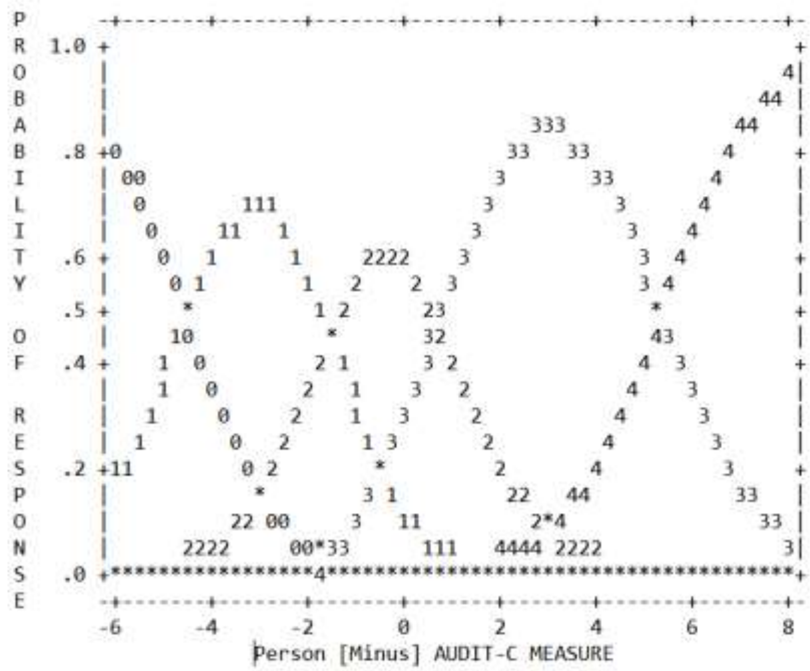

Figure 5. AUDIT-C rating scale use.

```
-----------------------------------------------------------------------------

MEASURE                | BOTTOM P=50%  | MEASURE       | TOP P=50%    MEASURE
 <more> ----- PERSON -+- ITEM        -+- ITEM        -+- ITEM         <rare>
   8                  +               +               +                8
                      |               |               |
                      |               |               | AUDIT-C-2
   7                . +               +               +                7
                      |               |               |
                      |               |               |
   6                  +               +               + AUDIT-C-3      6
                      |               |               |
                      |               |               |
   5                  +               +               +                5
                    . |               |               |
                      |               |               |
   4                  +               +               +                4
                      |               |               |
                      |               |               |
   3               ## +               +               +                3
                      |               |               |
                      |               |               | AUDIT-C-1
   2               .# +               + AUDIT-C-2     +                2
                      |               |               |
                      |               |               |
   1             .### +               +               +                1
                      |               | AUDIT-C-3     |
                      |               |               |
   0            .#### +               +               +                0
                      |               |               |
                      |               |               |
  -1                  +               +               +               -1
             ######## |               |               |
                      |               |               |
  -2                  +               +               +               -2
         .########### |               |               |
                      | AUDIT-C-2     |               |
  -3                  +               + AUDIT-C-1     +               -3
          .########## |               |               |
                      |               |               |
  -4                  + AUDIT-C-3     +               +               -4
           .######### |               |               |
                      |               |               |
  -5                  +               +               +               -5
                      |               |               |
                      |               |               |
  -6                  +               +               +               -6
            .######## |               |               |
                      |               |               |
  -7                  +               +               +               -7
                      |               |               |
                      | AUDIT-C-1 |   |               |
  -8      .########## +               +               +               -8
 <less> ----- PERSON -+- ITEM        -+- ITEM        -+- ITEM         <frequent>
EACH "#" IN THE PERSON COLUMN IS 7 PERSON: EACH "." IS 1 TO 6
```

Figure 6. Map of person and item for AUDIT-C.

Research Question Six

Hierarchical linear regression was used to investigate the relationships between the PHQ-4 and impulsivity questions and binge drinking items when controlling for demographic variables.

Hierarchical multiple regression was used to investigate the utility of different demographic characteristics, impulsivity, and depression to predict alcohol consumption and binge drinking. Based on the literature, several different models were examined using hierarchical multiple regression that employed different variables as the dependent variable. Prior to conducting a hierarchical multiple regression, the relevant assumptions of the statistical analysis for this study were tested. First, a sample size of 491 was deemed as adequate, given the predictors (Tabachnick & Fidell, 2007). Multicollinearity and singularity were deemed to be within accepted limits, as indicated by tolerance and VIF. Residual and scatterplots were reviewed, and assumptions of normality, linearity, and homoscedasticity were all satisfied. No outliers were found. Table 13 provides descriptive information about the variables and their relationships.

Table 13

Correlations, Means, Standard Deviations, and Alphas among the Variables

Variable	PHQ-4	Impulsivity	AUDIT-C Q1	AUDIT-C Q2	AUDIT-C Q3
PHQ-4 Person Position	1				
Impulsivity Person Position	.23*	1			
AUDIT-C Q1	.01	-.01	1		
AUDIT-C Q2	.02	.15*	.41*	1	
AUDIT-C Q3	-.00	.12*	.57*	.64*	1
Mean	-4.22	-2.05	1.86	.56	.85
SD	2.29	2.41	1.17	.79	.94
Skewness	1.19	.51	-.22	1.50	.89
Kurtosis	.65	.79	-.96	2.20	-.19

Note: $N = 491$. * $p < .01$

For the PHQ-4 and the four-item impulsivity scale, Rasch analysis indicated that person position measures could be used for the hierarchical regression. Correlations were calculated between the total score for each measure and the Rasch person position estimates. Person position should be used if there is a strong correlation, as it yields a more accurate analysis. Correlations between the Rasch person positions and the total score for each scale exceeded 0.8. Specifically, the PHQ-4 total score with the Rasch person position correlation was .99 (p<.001), and was 0.98 (p<.001) for the four-item impulsivity total score with the Rasch person position. The PHQ-4 Rasch person position and the four-item impulsivity person position were used for the hierarchical regression. Since Rasch scale analysis indicated substantial DIF in items for the AUDIT-C, the three AUDIT-C items were used as a separate dependent variable.

PHQ-4 Hierarchical Regressions

The first hierarchical multiple regression was conducted with the PHQ-4 person position as the dependent variable. Age (categorized as either 22 years and under or 23 years and older), gender (female or male), and ethnicity (characterized as white or non-white) were entered as block one of the regression, with the impulsivity person position entered as block two. Collinearity statistics did not indicate multicollinearity. Scatterplots were examined for outliers, normality, linearity, and homoscedasticity. Regression statistics are in Table 14.

The hierarchical regression revealed that at block one, gender, ethnicity, and age accounted for 1.4% of the variation in PHQ-4 person position, and was a non-significant model. Introducing the impulsivity variable explained an additional 5.1% when controlling for gender, ethnicity, and age, and the full model was statistically significant: $p < .001$. At block two, the model was statistically significant as a whole: $F(4, 486) = 8.38$; $p < .001$. The best predictor of PHQ-4 score was impulsivity person position (Beta = .230; $p < .001$) in the full model, with age as the best, though non-significant predictor (Beta = 0.08; $p = 0.73$). This indicates that the addition of the impulsivity questions aided in identifying individuals that are depressed that demographic variables alone might not have detected. Specifically, a higher impulsivity person position correlated with a higher level of depression.

Table 14

Summary of Hierarchical Regression Model for Variables Predicting PHQ-4

Step	R	R2	ΔR^2	*B*	SE	Beta	*t*	*p*
Step 1	.12	.01	.01					.08
Gender				.44	.22	.09	2.01	<.05
Age				.24	.22	.05	1.09	.28
Ethnicity				.30	.256	.05	1.19	.23
Step 2	.25	.06	.05					<.001
Gender				.26	.21	.06	1.23	.22
Age				.40	.22	.08	1.79	.07
Ethnicity				.30	.25	.05	1.19	.24
Impulsivity				.22	.04	.23	5.12	<.001

Note: N = 491

AUDIT-C Hierarchical Regressions

Since the Rasch analysis indicated DIF on the AUDIT-C items, three different regressions were run with each AUDIT-C item as the dependent variable (DV). Block one included demographic items, and the second block was the Rasch PHQ-4 and impulsivity person positions.

The second hierarchical multiple regression was conducted with AUDIT-C question 1 as the dependent variable. AUDIT-C question 1 asks "How often do you have a drink containing alcohol?" Age (categorized as either 22 years and under or 23 years and older), gender (female or male), and ethnicity (characterized as white or non-white) were entered at block one of the regression, with the Rasch PHQ-4 and impulsivity person positions entered at block two. Collinearity statistics did not indicate multicollinearity. Scatterplots were examined for outliers, normality, linearity, and homoscedasticity. Regression statistics are reported in Table 15. The hierarchical regression revealed that at block one, gender, ethnicity, and age, accounted for 5.3% of the variation in the question 1 score of the AUDIT-C. Introducing the PHQ-4 and impulsivity variables explained a slight increase of .2% when controlling for gender, ethnicity, and age. In the first model with the three predictors of gender, ethnicity, and age entered, the model was found to be statistically significant: $F(3, 487) = 9.1$; $p < .001$. The ANOVA table indicated that the second model was also significant: $F(5, 485) = 5.647$; $p < .001$. The best predictor of AUDIT-C question 1 was ethnicity (Beta = -.179; $p < .001$) in the full model. Gender was also a statistically significant predictor of AUDIT-C question 1 in the full model (Beta = -.15; $p < .001$). This indicates that male drinking frequency was significantly more than females, and that the drinking frequency of white individuals was significantly higher than non-whites.

Table 15

Summary of Hierarchical Regression Model for Variables Predicting AUDIT-C Q1

Step	**R**	**R2**	**ΔR^2**	***B***	**SE**	**Beta**	***t***	***p***
Step 1	.23	.05	.05					<.001
Gender				.37	.08	.15	3.39	<.01
Age				-.02	.11	-.01	-.15	.88
Ethnicity				-.52	.13	-.180	-4.04	<.001
Step 2	.24	.06	.002					<.001
Gender				.38	.11	.15	3.45	<.01
Age				-.03	.11	-.01	-.30	.76
Ethnicity				-.52	.13	-.18	-4.05	<.001
PHQ-4				.01	.02	.02	.45	.66
Impulsivity				-.02	.02	-.04	-.95	.34

Note: $N = 491$.

The third hierarchical multiple regression was conducted with question 2 of the AUDIT-C as the dependent variable. AUDIT-C question 2 asks "How many drinks containing alcohol do you have on a typical day when you are drinking?" Age, (categorized as either 22 years and under or 23 years and older), gender (female or male), and ethnicity (characterized as white or non-white) were entered at block one of the regression with the PHQ-4, and impulsivity Rasch person position was entered at block two. Collinearity statistics did not indicate multicollinearity. Scatterplots were examined for outliers, normality, linearity, and homoscedasticity. Regression variables are reported in Table 16. The hierarchical regression revealed that at block one (gender, ethnicity, and age) accounted for 14.6% of the variation of the AUDIT-C question 2 score. Introducing the impulsivity variable explained an additional .6% when controlled for gender, ethnicity, and age. The first model was statistically significant: $F(3, 487) = 27.74$; $p < .001$. The second model was also significant: $F(5, 485) = 17.33$; $p < .001$. The best predictor of question 2 of the AUDIT-C was gender (Beta = .260; $p < .001$) in the full model and also in the first-level model (Beta = .272; $p < .001$); though, all three demographic variables were significant: $p < .05$. This indicates that for white males aged 22 and under, drinking amount is significantly more than it was for females, and the drinking amounts of white individuals was significantly more than non-whites. Also, the age group 22 years and under drinks significantly more than the 23 years and older age group. The PHQ-4 and impulsivity person positions were not significant in the full model.

Table 16

Summary of Hierarchical Regression Model for Variables Predicting AUDIT-C Q2

Step	**R**	**R2**	**Δ*R*²**	***B***	**SE**	**Beta**	***t***	***p***
Step 1	.38	.15	.15					<.001
Gender				.45	.07	.27	6.48	<.001
Age				-.42	.07	-.25	-5.87	<.001
Ethnicity				-.17	.08	-.09	-2.06	<.05
Step 2	.39	.15	.01					<.001
Gender				.43	.07	.26	6.12	<.001
Age				-.40	.07	-.24	-5.56	<.001
Ethnicity				-.170	.08	-.09	-2.07	.04
PHQ-4				-.01	.02	-.01	-.08	.94
Impulsivity				.03	.01	.08	1.76	.08

Note: N = 491

The fourth hierarchical multiple regression was conducted with question 3 of the AUDIT-C as the dependent variable. Question 3 of the AUDIT-C is aimed at binge drinking, and asks “How often do you have six or more drinks on one occasion?” Age, (categorized as either 22 years and under or 23 years and older), gender (female or male), and ethnicity (characterized as white or non-white) were entered at block one of the regression with the PHQ-4, and impulsivity Rasch person position was entered at block two. Collinearity statistics did not indicate multicollinearity. The P-P and scatterplots were examined for outliers, normality, linearity, and homoscedasticity. Regression statistics are outlined in Table 17. The hierarchical regression revealed that at block one, gender, ethnicity, and age, accounted for 18.2% of the variation in question 3 of the AUDIT-C score. Introducing the impulsivity variables explained only an additional .3%

when controlling for gender, ethnicity, and age. In the first step, the three predictors were entered, and it was found that gender, ethnicity, and age were significant: $F(3, 487) =36.06$; $p < .001$. The ANOVA table indicated that the model as a whole was significant: $F(5, 485) =21.959$; $p < .001$. The best predictor of question 3 of the AUDIT-C was gender (Beta = .372; $p < .001$) in the full model and also in the first level-model (Beta = .377; $p < .001$). However, all three demographic variables were significant predictors: $p < .001$. This indicates that for white males aged 22 years and under, binge drinking was significantly higher than it was for females, and also that the drinking amount of white individuals was significantly more than it was for non-whites. The 22 years and under age group also drinks significantly more than the 23 years and older age group. The PHQ-4 and impulsivity person positions were not significant in the full model.

Table 17

Summary of Hierarchical Regression Model for Variables Predicting AUDIT-C Q3

Step	**R**	**R2**	**ΔR^2**	***B***	**SE**	**Beta**	***t***	***p***
Step 1	.43	.18	.18					<.001
Gender				.75	.08	.38	9.20	<.001
Age				-.26	.08	-.13	-3.07	<.01
Ethnicity				-.35	.10	-.15	-3.64	<.001
Step 2	.43	.19	.01					<.001
Gender				.73	.08	.37	8.94	<.001
Age				-.24	.09	-.12	-2.82	<.01
Ethnicity				-.35	.10	-.15	-3.60	<.001
PHQ-4				-.01	.02	-.03	-.75	.45
Impulsivity				.02	.02	.05	1.21	.23

Note: $N = 491$

Research Question Seven

LCA analysis will be used to determine if there are undetermined classes present.

The latent factor structure of the brief screening instrument was assessed by use of exploratory latent class analysis (LCA) techniques with *Mplus* (Version 7.11; Muthen & Muthen, 2012). The results were evaluated using Wang and Wang's (2012) three-step approach of determining the optimal number of latent classes, evaluating the quality of latent class membership, and defining the latent classes. Responses from the PHQ-4 Rasch person position, the three individual AUDIT-C questions, and the four-item impulsivity measure Rasch person positon were entered into the LCA model, as initial research showed that there were three theoretical constructs: depression, alcohol use/abuse, and impulsivity. Prior to analysis, the data was examined to ensure that sufficient values were in each cell of the contingency table. All entered items were treated as ordered, categorical variables in the model. The *Mplus* number of iterations was initially set to 1,000, and default starting variables were used for this analysis. However, after the first analysis failed to replicate the best log of likely values, the starts were increased to 2,000 with more random starting position. A typical *Mplus* input file specification for this analysis is presented in Appendix F.

The optimal number of classes was determined by analyzing the fit of a series of increasing class number models by comparing the fit statistics and information criterion indices for each of the models, which ranged from one to six latent classes (see Table 18).

Table 18

LCA Model Comparison

Statistic/Index	**1-Class**	**2-Class**	**3-Class**	**4-Class**	**5-Class**
LMR LRT *p*-value	N/A	<0.001	0.0025	0.0538	0.3240
ALMR LRT *p*-value	N/A	<0.001	0.0027	0.0569	0.3293
BLRT LRT *p*-value	N/A	<0.001	<0.001	<0.001	<0.001
AIC	10264.283	9760.991	9423.176	9399.550	9430.145
BIC	10318.837	9853.313	9553.266	9567.407	9635.771
ABIC	10277.575	9783.485	9454.873	9440.448	9480.246
Entropy	N/A	.904	.999	.971	.888

Note: LMR LRT = Lo-Mendel-Rubin Likelihood Ration Test; ALRM LRT = Adjusted Lo-Mendell-Rubin Likelihood Ratio Test; BLRT = Bootstrap Likelihood Ration Test

The optimal number of classes was determined by analyzing fit by increasing the class number by one and comparing the fit statistics. The fit statistics and criterion indices for the models, ranging from one to five latent classes, are displayed above in Table 18. Both the LMR LR test (p=.0538) and the ALMR LR test (p=.0569) were statistically non-significant in the four-class model, so the three-class model was determined to be the optimal number of classes based on model fit. Further support for the three-class model was supported by the BIC, which decreased through the three-class model, but increased with a four-class model.

While the number of individuals into a latent class was not definitely determined, individuals were assigned into a latent class based on the highest probability for the class. The class counts based on estimated posterior probabilities for each individual assigned to a class are given in Table 19. Table 19 shows that 221 individuals were assigned to class one, 163 individuals were assigned to class two, and 107 individuals were assigned to class three, yielding adequate sizes and samples among the classes.

Table 19

Final Latent Class Counts and Proportions

Classes	Counts	Proportions
1	221	45.01%
2	163	33.20%
3	107	21.92%

As shown in Table 20, the average latent class probability of correct class membership for individuals assigned to class one was ~1.000, while the probability of misclassification was < .001. For the second class, the probability of correct membership was ~1.000, with the probability for misclassification was < .001. The third class resulted in a probability of correct membership of ~1.000, with the probability of misclassification being < .001. These average latent class probabilities for most likely latent class membership well exceeded Nagin's (2005) criterion for minimum acceptable class membership of 0.7 for all groups.

Table 20

Average Latent Class Probabilities for Most Likely Latent Class Membership

Classes	Probability of Class 1 Membership	Probability of Class 2 Membership	Probability of Class 3 Membership
1	~1.000	<.001	<.001
2	<.001	~1.000	<.001
3	<.001	<.001	~1.000

Another criterion to summarize posterior misclassification is based on entropy, a single value summary of the degree of uncertainty in the model, scaled such that large values indicate less classification error (Collins & Lanza, 2010). The entropy statistic for the three-class model was .855, which is considered a high value (Clark, 2010). Thus, it can be concluded that latent class membership was satisfactory.

As shown in Table 21 and Figure 7, class one was comprised mostly female participants (72.6% probability) and the oldest participants (41.8% probability of being 23 years and over). Class one had the highest number of non-white individuals, and reported drinking less than any of the other classes. The class had the highest level of depression, and was the least impulsive of all the groups. Class two was mostly female (62.7% probability) and was the youngest class (83.7% probability of being 22 years and younger). This group had the lowest level of overall depression compared to the other two classes. The class drinks, but does not binge drink, and is more impulsive than the population mean. Class three was mostly male (28.3% probability) and, as a whole, was the youngest group (88.7% probability of being 22 years or younger). This group drank the most and binge drank at high levels. Class three was the most impulsive of all groups.

Table 21

Descriptive Statistics of the Latent Classes

	All Classes N=491	**Class 1 N=221 (45.01%)**	**Class 2 N= 163 (33.20%)**	**Class 3 N=107 (21.92%)**
Gender	64.8% Female 35.2% Male	72.6% Female 27.4% Male	62.7% Female 37.3% Male	28.3% Female 71.7% Male
Age	69.5% $\leq$ 22 30.5% $\geq$ 23	58.2% $\leq$ 22 41.8% $\geq$ 23	83.7% $\leq$ 22 16.3% $\geq$ 23	88.7% $\leq$ 22 11.3% $\geq$ 23r
Ethnicity	79.4% white 20.6% non-white	76.5% white 23.5% non-white	83.0% white 17.0% non-white	84.9% white 15.1% non-white
PHQ-4 Mean	-4.216	-4.351*	-3.976*	-4.301*
Impulsivity Mean	-2.052	-2.260*	-2.138*	-1.491*
AUDIT-C Q1 Mean	1.86	1.181*	2.129*	2.842*
AUDIT-C Q2 Mean	0.56	0.154*	.583*	1.384*
AUDIT-C Q3 Mean	0.85	0.000	0.051*	2.374*

Note: Means are reported as coefficients standardized using the variance of the continuous latent variable. *p<.05

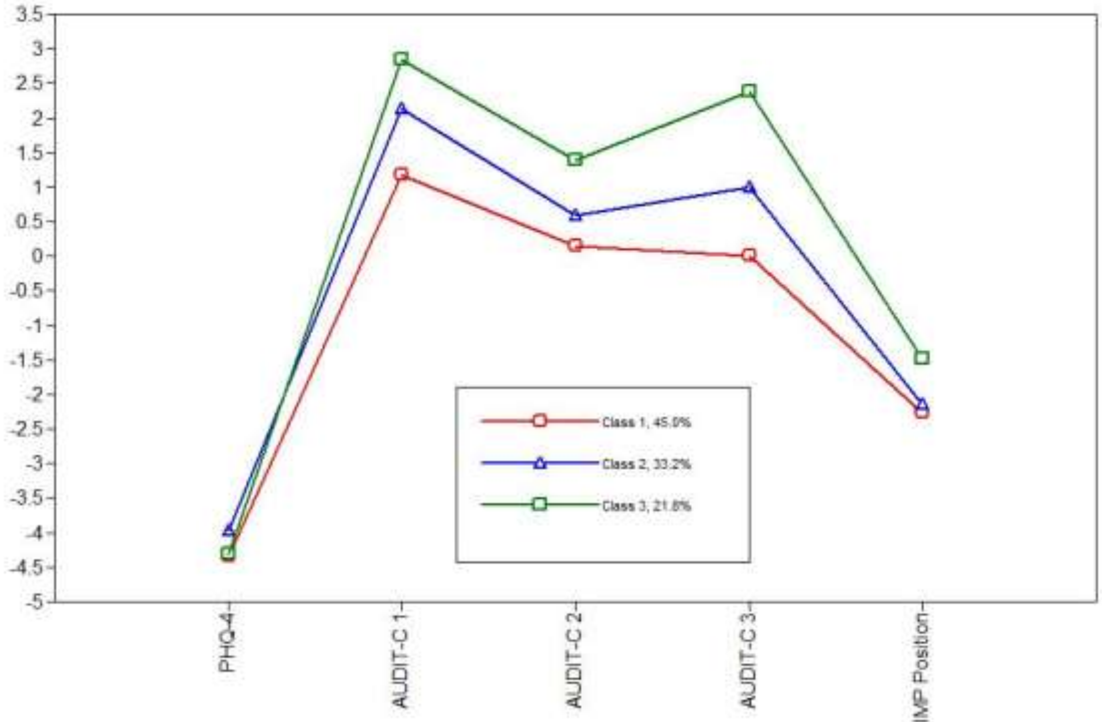

Figure 7. Latent Classes of Sample.

CHAPTER FOUR: DISCUSSION

This chapter presents a summary of the study, major findings according to each research question, integration of results of the four analysis techniques, limitations of the research study, and recommendations for further study.

Summary and Major Findings

This study introduced the theoretical need for a brief screening instrument that could be used to screen for depression and alcohol misuse in a college-aged population. This study reviewed literature on brief screening instruments that are currently in use in primary care medical settings for this purpose. The two screening instruments chosen for this study, which are in wide use, were the PHQ-4 for depression and the AUDIT-C for excessive alcohol consumption. While both instruments are extensively used, they have undergone little analytical review.

Researchers have been building a body of evidence that shows a relationship between impulsivity and depression (in the case of suicidal ideation) and with alcohol (in relation to excessive drinking). Research in this area has also focused on identifying different facets of impulsivity and relating them to harmful behaviors. One instrument used more broadly to diagnose impulsivity is the UPPS-P measure, which has 59 questions and does not lend itself to brief screening settings. The UPPS-P identifies five facets of impulsivity: Urgency (lack of); Premeditation (lack of); Perseverance; Sensation Seeking; and Positive Urgency. Shorter brief screening instruments are currently being developed to identify these facets and also to identify the facets most closely related to harmful and depressive behaviors and alcohol misuse. The focus of this study was to develop a brief screening measure to better help identify depression level and excessive alcohol consumption—an instrument that included several items from the UPPS-P. This study also articulated the need for additional psychometric analyses and substantive interpretation to strengthen the rigor of the PHQ-4 and AUDIT-C instruments in a college-aged population.

Chapter 2 described the application of exploratory factor analysis, item response theory, hierarchical regression, and latent class theory in assessing the latent factor structure of this new instrument and the relationships among scales and demographic variables.

Exploratory factor analysis (EFA) was used to provide evidence in order to predict factor structure based on *a priori* hypotheses, to provide statistical criteria regarding the underlying factors, to test and compare alternative models to the data, and to determine the dimensionality of the measurement. As anticipated, the structure comprised three factors that were the underlying instruments (the PHQ-4 and the AUDIT-C) and the eight items that were from the UPPS-P measure.

Using the factors from the EFA, the 15-item measure used in this study was analyzed using the nonlinear approach from item response theory. Rasch analyses were used to explore the dimensionality of the full 15-item scale and to investigate the dimensionality of the three factors identified by EFA. The 15-item measure was multidimensional (as expected), but the three factors (PHQ-4, AUDIT-C, and impulsivity) were found to be acceptably unidimensional. One item was dropped during the EFA as it was found to have inadequate fit with any of the three factors. To have a unidimensional impulsivity measure, four items were eliminated from the eight impulsivity questions that formed the basis of the new impulsivity measure. Of note was that the AUDIT-C showed differential item functioning for both gender and age for two of the three items; therefore, it was not treated as a scale, with further analyses using the individual items of the AUDIT-C rather than the Rasch person position. Further analysis of the PHQ-4 and the impulsivity measure used Rasch person position logits.

These techniques were followed by hierarchical regression (HR). Demographics were entered as the first level of the HR, with the dependent variable being either the PHQ-4 Rasch person position or the individual AUDIT-C items. The second level was the Rasch person position of the new impulsivity measure, with the addition of the PHQ-4 Rasch person position for the AUDIT-C questions. This analysis was beneficial, as it confirmed differences between groups based on gender, age, and ethnicity. The addition of the impulsivity items to the model resulted in a significant contribution to the PHQ-4. This is important to note because, other than the addition of more depression questions that would make the PHQ-4 a more diagnostic tool rather than a screening tool, the impulsivity measure could be an important contribution to explaining variance in PHQ-4 scores. There were no significant contributions to predicting the alcohol item variance after adding the PHQ-4 and impulsivity Rasch person positions. This was contrary to what was expected, as the literature has shown showed a consistent relationship between facets of impulsivity and problematic drinking behaviors.

Latent class analysis (LCA) permitted classifying mixtures of individuals into subpopulations based on their responses to the resulting measure. The analyses uncovered the number of underlying subpopulations. Identifying the latent class models of the study complemented the other dimensional approaches of the structure assessment, and helped to identify subpopulations that could lead to future research and interventions. Three classes were discovered, with the most interesting being a younger, mostly white, and male class (21.9% of the population) that had extremely high levels of alcohol consumption. Another group was a younger, mostly female, white class (45% of the population) that had a higher level of reported depression, but did not engage in higher levels of alcohol consumption. The final group, which represented 33.2% of the population, fit in between the other two groups, comprised mostly younger, white females who were the least depressed group that drank regularly, but did not engage in binge drinking.

Major Findings by Research Question

This section discusses the major findings of the study based on each research question. Interpretation of the results is based on the literature review provided earlier.

Research Question One

Exploratory factor analysis was used to determine if the new scale has a unidimensional structure. It was hypothesized that the structure is multidimensional.

This research question asked if the structure was unidimensional or multidimensional. Because the scale was made up of two validated instruments—the PHQ-4 and the AUDIT-C—with the addition of a series of eight impulsivity questions that were intended to improve the identification of either depression or alcohol misuse, it was expected the scale would be multidimensional. Principal axis factoring was used for the factor analysis, with orthogonal rotation (Varimax). Several different models were examined, including principal components analysis with oblique rotation, but all yielded similar models: as expected, all models produced a multidimensional structure. The three-factor solution, which explained 57.32% of the variance, was preferred because of support from parallel analysis. One question, UPPS-P question 27 was dropped, as it did not have a factor loading above .30 on the three retained factors. The final 14-item factor solution was produced using PAF with oblique rotation. The final structure had four items that loaded on factor one (the PHQ-4) and three items that loaded on factor two (the AUDIT-C); the remaining questions were the UPPS-P questions. This structure corresponded to the underlying theory used to develop the instrument.

Research Question Two

This question asked if analysis of the 15-item measure using Rasch IRT would support a unidimensional structure. Dimensionality was examined for the measure by analyzing the overall fit, examining a principal components analysis of residuals, and reviewing individual item fit. If unidimensional, differential item function was examined.

Rasch IRT analysis, using *Winsteps* (Version 3.81.0), was employed to determine if the 15-item measure was unidimensional. Dimensionality was explored by analyzing overall and individual item fit and principal components analysis of residuals. While the 15-item measure had good fit initially for both person and items, a deeper review of the items showed severe individual item misfit (infit mean square ranged from 1.94 to .38, and mean square outfit ranged from 2.16 to .37). Principal components analysis of residuals also indicated multidimensionality with the first contrast eigenvalue at 3.2, which is higher than the recommended 2.0 or less (Linacre, 2010). Based on this, it was concluded that the 15-item scale was multidimensional; therefore, further Rasch analysis was not performed.

Research Question Three

Rasch analysis of the PHQ-4 examined the dimensional structure in a college-aged population and the presence of differential item functioning by gender, year in college, and ethnicity.

Rasch analysis using *Winsteps* (Version 3.81.0) was employed to determine if the PHQ-4 four-item measure was unidimensional. Dimensionality was explored by analyzing overall and individual item fit and principal components analysis of residuals. The measure had good fit for both person and items, and a Rasch principal components analysis of residuals yielded a first contrast eigenvalue of 2.1, which is slightly above the value of 2.0 to be considered unidimensional (Linacre, 2010). The measure displayed low reliability with person separation, being .97 (1.24 when removing extreme persons). The low separation is problematic, but is also indicative of a short item screening instrument. Since the majority of the population reported low levels of depression, the scale did not have a high range of responses. To be most useful, instruments should have separation of at least 2.0 (Linacre, 2010). However, practitioners find the PHQ-4 valuable as a screening tool. Differential item function was also examined for gender, age, and ethnicity in order to investigate invariance. There was DIF for gender on question 3, which is one of the depression questions. This is consistent with research indicating that it is easier for females to agree on depression questions than for males. A possible remedy for this would be to add more depression questions to balance the DIF item.

However, with a brief screen that is intended to be completed quickly, this adds complexity to the scoring by the practitioner, and also time that could be used by the provider to further explore diagnosis. The DIF on gender should be investigated further but, for the current study, the PHQ-4 logit person position was used across genders.

Targeting and scale use for the PHQ-4 were also an issue. Based on the distribution, it appears that only a small percentage of this population was measured well by the PHQ-4, with the majority of the sample falling below -5.0. This should be expected with a depression measure, especially when the majority of the population has an absence of depression. The item order and scale used was as anticipated, with no inversions in the step structure.

Research Question Four

> **Rasch analysis of the impulsivity factor identified by EFA was used to examine the dimensional structure in a college-aged population. Dimensionality was examined. If unidimensional, differential item function was examined.**

Rasch analysis was employed to determine if the impulsivity measure identified by EFA displayed a unidimensional structure. The first contrast eigenvalue was 2.9, which was well above the value of 2.0 needed to be considered unidimensional (Linacre, 2010). Individual item fit was examined, and showed several poorly fitting items. Four items (questions 1, 2, 5, and 8) were removed by eliminating the underfitting items one at a time. One of the items removed was the poorly fitting items identified in the EFA analysis. The remaining four-item measure displayed good fit, and also resulted in the near adequate person separation, at 1.67. The first contrast eigenvalue was 1.9, which was below the value of 2.0 needed to be considered unidimensional (Linacre, 2010). Overall fit was considered adequate. Differential item function was also examined for gender, age, and ethnicity in order to investigate invariance. There was DIF for gender on question 1 (p<.05) and DIF for age on questions 1 and 4 (p<.01), but the DIF contrast was not large. The DIF on gender and age should be investigated further by creating a longer instrument that has questions from the same facets from which the UPPS-P questions for this study came. There was no DIF on ethnicity. For the current study, the impulsivity logit person position was used because the impulsivity scale was determined to be useful as a measure.

Targeting and scale use for the impulsivity measure were good, suggesting that impulsivity was measured well in this sample. A number of individuals fell below a zero score; however, this can be explained: The majority of the individuals taking the items were female, and they tend to be less impulsive than men. The item order and scale use was as anticipated, with no inversions in the step structure.

Research Question Five

Rasch analysis of the AUDIT-C was used to examine the dimensional structure in a college-aged population. Dimensionality was examined. If unidimensional, differential item function was examined.

Rasch IRT analysis was employed to determine if the AUDIT-C displayed a unidimensional structure. The measure had good fit for both person and items, with a first contrast eigenvalue of 2.4, which was above the value of 2.0 needed to be considered unidimensional (Linacre, 2010). However, individual items displayed good person and item fit, with a person separation of 1.84. Invariance was measured by examining differential item function for gender, age, and ethnicity in order to investigate invariance. There was significant and substantial DIF for gender on questions 1 and 3 (p<.001), and DIF for age on questions 1 and 3 (p<.001). As DIF was both statistically significant and substantial, it was concluded that the AUDIT-C could not reasonably be treated as an invariant measure. The AUDIT-C was not treated as a scale; therefore, further analyses used items in the measure individually. The DIF on gender and age should be investigated further, and is problematic. Again, the problem of a short screening instrument is apparent. One possible solution would be to add or modify items that could help eliminate DIF. This is a widely used instrument, and prior research shows differences in achieving different thresholds for different populations or genders. It is suggested that further research be undertaken to determine the source and nature of this DIF.

Scale use for the AUDIT-C measure was good. Item order and scale use was as anticipated, with no inversions in the step structure. Targeting appeared to be acceptable, with the majority below a score of zero; however this is somewhat questionable given that this is a college-aged population, with a large percentage of individuals reporting moderate to heavy alcohol consumption.

Research Question Six

Hierarchical linear regression was used to investigate the relationships between the PHQ-4, impulsivity questions, and binge drinking items when controlling for demographic variables.

Hierarchical linear regression was used to investigate the utility of different demographic characteristics, impulsivity, and depression in predicting alcohol consumption and binge drinking. PHQ-4 and impulsivity Rasch person positions were used for the regression, but because the AUDIT-C displayed significant DIF, individual items were used for the regressions. Four different regressions were performed, with the demographics of gender, age, and ethnicity being the first block in all of the models.

The first hierarchical regression used the PHQ-4 person position as the dependent variable and the impulsivity person position as the predictor. This was the most surprising model of all the regressions. While the first level model was not significant, and only explained 1.4% of the variance in the model, when impulsivity was added at the second level, an additional 5.1% of the variance was explained, and the model as a whole was statistically significant. This indicates that the addition of impulsivity aids in identifying individuals that the PHQ-4 alone might not detect. However, the total variance explained was low.

The second through fourth hierarchical models used questions 1 through 3 of the AUDIT-C instrument, with the PHQ-4 and impulsivity person position in block two. All of these models were statistically significant at block 1 and in the full model. However, neither the PHQ-4 nor the impulsivity measure was significant at the second level; they also did not explain worthwhile additional variance in the full model. As noted in the literature review, alcohol consumption and binge drinking are the highest in young, white males, and the regression at the first block confirmed this. Age and gender were statistically significant in the two AUDIT-C consumption questions (questions 2 and 3; $p<.01$). However, what is noteworthy is that the demographics explained only 14.6% of the variance for AUDIT-C question 2, and 18.2% of the variance in question 3. It was expected that these demographics would have had a stronger contribution in explaining the variance. This casts more doubt on the validity of the AUDIT-C items as adequate measures for alcohol consumption. Finally, it should be noted that the scoring for the AUDIT-C has five response options that increase in difficulty (i.e., higher scores indicate more abusive drinking patterns). Question 1, which asks about the frequency of drinking times per month, had a higher mean than did the other two questions. A recommendation would be that if the frequency of drinking is low, the remaining items should not be completed. The intent is to capture the problematic drinking of individuals who actually drink. The AUDIT-C is presented as a validated measure with world-wide usage, and the results of the present study cast doubt upon that usage. Further study with a larger sample is needed in this area in order to investigate the AUDIT-C and its value as a scale.

Additional regressions not reported in the results were run by the researcher, with individual impulsivity items from the measure as the second block for the AUDIT-C items (as the dependent variable) or the PHQ-4 person position as the dependent variable. For the AUDIT-C items, the UPPS-P question 50—which asks, "When I am really excited, I tend not to think on the consequences of my actions"—explained a significant amount of the variance: approximately 2% when added in block two of the hierarchical model ($p < .01$). When other individual items from the UPPS-P that were discarded during the development of the impulsivity scale were added into the second block of the regression, question 48 (a Premeditation (lack of) question that asks, "I usually think carefully before doing anything") and question 29 (a Negative Urgency question that asks, "When I am upset, I often act without thinking") had significant contributions in explaining model variance—approximately 3% ($p < .001$) over block one (demographics) to 21.7%--when these three items were added into block two. UPPS-P questions 53 and

29 have been linked to binge drinking, and question 48 has a more theoretical relationship to suicidal ideation and bipolar disorder.

For the PHQ-4 person position as the dependent variable, examining the individual impulsivity scale revealed items identified by UPPS-P question 53, which asks "I tend to act without thinking when I am really excited," and explained nearly 2% or more variance in the regression model, where PHQ-4 was the DV, rather than using the four-item Impulsivity measure person positon. When examining other UPPS-P questions that were discarded during the development of the impulsivity measure, it was revealed that UPPS-P question 48, a Premeditation (lack of) question that asks, "I usually think carefully before doing anything," and UPPS-P 29, a Negative Urgency question that asks "When I am upset, I often act without thinking," increased the variance explained by the measure from 1.4% to 12.9%, and provided the best combination in explaining model variance.

Clearly, more research in this area is needed, and further study should be pursued in this area in order to develop brief depression and alcohol scales that include impulsivity items. Question selection from the full 59-item UPPS-P (specifically from the Premeditation and the Negative Urgency facets), displayed significant contribution ($p<.01$).

Research Question Seven

LCA analysis was used to determine if there are undetermined classes present.

Latent factor analysis was performed to assess the factor structure of the new 11-item measure. The PHQ-4 Rasch person position, the impulsivity Rasch person position, and the individual AUDIT-C items were used for this analysis. After testing five different models, a three-class solution was determined to have the best fit. The 491 respondents were identified as follows: Class 1, with 221 individuals (45%); Class 2, with 163 individuals (33%); and Class 3, with 107 individuals (22%). Class 1 was mostly female (72.6% probability), and was the oldest class (58.2% probability of being 22 years old and under). It had the highest number of non-white individuals, and drank less than any of the other classes. This class had the highest level of depression, and was the least impulsive of all the groups. Class 2 was mostly female (62.7% probability), and was the youngest class (83.7% probability of being 22 years old and younger). This group had the lowest level of overall depression compared to the other two classes, drank moderately but did not binge drink, and was a bit more impulsive than the mean. Class 3 was mostly male (28.3% probability) and, as a whole, was the youngest group (88.7% probability of being 22 and younger). This group drank the most and binge drank at high levels. Class three was the most impulsive of all groups.

Identification of these groups can assist future research in identifying targeted interventions for specific groups (such as Class 3) for more intense alcohol education and prevention efforts.

Reliability versus Utility in a Brief Screening Measure

A key issue that arose repeatedly in this study was difficulty achieving adequate person separation. Rasch person separation determines the reliability index. In all of the models investigated, person separation was low (< 2; with person reliability, <0.8) and, in some cases, extremely low. This implies that the instruments might not have been sensitive enough to distinguish between high and low performers. A typical solution to this issue would be to add more items; however, this would be problematic since the intention of these instruments was to have a brief screening tool that could quickly identify issues that indicated a need for further diagnostic testing. Those who develop measures desire high reliability to ensure that there is a high probability that the instrument consistently measures what was intended. For high reliability, a variable sample and low measurement error is needed. For high person reliability, a sample with a diverse ability range, and an instrument with many items, is needed. The three brief screening instruments in this study screen for traits that typically have a modest or low percentage of occurrence in the population. This is analogous to having a limited ability sample. For example, it is common to find low levels of depression in a population with moderate to high levels of depression that affects a small percentage of the population. Therefore, when measured on a brief screening instrument for depression, the “difficulty” does not have a wide range.

By their nature, screening measures are intended to only screen for problems that, typically, occur infrequently. In many ways, screening measures are meant to identify outliers—people who are moderately or severely depressed in a general population where depression typically is not common.

A second complicating issue is that brief screening instruments are also designed to be short; they are not intended to be diagnostic measures, which would contain more items and could be designed to have a wider ability range and perform better as a scale. There is a tradeoff that occurs with these types of instruments: If the focus is on developing a good measure, screening instruments would likely not work well for the brief screening needed in many practical applications. High reliability or separation with a four-item instrument is unlikely to be found, especially for complex constructs such as depression. Complicating the issue further is that scoring is intended to be simple. This creates another limitation, making it difficult for the instrument designer to increase the number of responses to an item in order to increase the variance in person position. In summary, it is difficult to develop a brief screening instrument that can demonstrate high reliability because, at its basis, what is needed for higher reliability as a measure is contrary to the design and nature of brief screening tools.

However, even with lower reliability, these instruments are extremely useful clinically since they screen for, and identify, many individuals who might not be otherwise identified. This is a tradeoff of reliability for utility. Table 22 provides reliability coefficients found in the literature and the internal consistency reliability estimated in the present study.

Table 22

Summary of Reliability for Measures Used

Measure	**Published Reliability**	**Reliability in Current Study**
PHQ-4	.82*	.85
AUDIT-C	.58**	.77
UPPS-P	.94***	N/A
UPPS-P Short	.70-.84****	N/A
4-item Impulsivity Scale	N/A	.80

Note: *Lowe et al., 2010, Simon et al., 2013, **, Verdejo-Garcia et al., 2010***, Billieux et al., 2012****

Summary of the Research

The purpose of this study was to develop and analyze a brief screening tool that captures depression and alcohol misuse, specifically binge drinking, since there is no instrument currently available to address this need in primary care settings. This research focused on improving detection by adding selected impulsivity items that previous research had shown to be identified with these issues from the 59-item UPPS-P into the existing and previously validated PHQ-4 and AUDIT-C measures. The goal of the research was to enhance the ability of this new brief instrument to better assist in identifying these problems. Brief screening tools are becoming critical in primary care to identify underlying issues that can be addressed in treatment and prevention. Currently, there is a lack of such measures that have been rigorously evaluated for widespread use.

The resulting measure piloted for this research was a 15-item measure that was administered to 491 college-aged individuals. Using the results of an initial exploratory factor analysis on the 15-item measure, three factors were identified: depression, alcohol use, and impulsivity. Since Rasch analysis of the full 15-item indicated multidimensionality, the three smaller scales (the PHQ-4, the AUDIT-C, and the eight impulsivity questions) were examined for unidimensionality, contribution to model fit, and explanation of contribution to measure variance, and were used to identify any undiscovered classes within the population.

Prior investigation of the PHQ-4, a relatively new instrument, using confirmatory factor analysis indicated acceptable unidimensional fit (Löwe, 2008). Further research on the PHQ-4 indicated that the results for this instrument were similar to earlier research on the longer PHQ-9 (Löwe, 2010). In the present study, the PHQ-4 was examined using

Rasch analysis, and was found to be unidimensional. Person separation was slightly lower than desired, and there was also DIF on gender. Targeting and scale use were also an issue, with only a small percentage of the population being measured well; however, scale and item fit were adequate and for this population. Overall, in this research, the PHQ-4 functioned as expected, and appears to be a useful measure for detecting depression in a college-aged population.

The three-item AUDIT-C, which is a subset of the ten-item AUDIT, had been shown to have nearly identical psychometric properties as its longer version (Menses-Gaya et al., 2010). While it was unclear if a Rasch analysis of the AUDIT-C had been previously conducted, earlier research indicated some issues with the measure as a scale, resulting in varied scoring based on gender, ethnicity, and age (Aalto, 2009; Dawson, 2012; Bradley et al., 2007; Graham, 2007). This study found that the AUDIT-C had a significant and substantial DIF by gender and age, and should not be treated as a measure in this setting. While there is considerable utility in a brief screening setting for identification of problematic consumption, as far as using this as a measure to compare scores between individuals, the AUDIT-C was inadequate. This is an important issue because the AUDIT-C is one of the mostly widely used consumption measures worldwide. Since the three-item AUDIT-C did not function as a measure in this population, and has been previously shown to have varied reliably in other populations, future research should focus on scale analysis of the full AUDIT to determine if its performance is adequate in other populations. The full AUDIT might function adequately as a measure while the 3-item brief scale does not.

Impulsivity items were selected from the 59-item UPPS-P based on earlier research (Whiteside & Lynam, 2001; Cyders et al., 2007). Item selection was enhanced by using factor loadings for the impulsivity facets that were most closely associated with suicidal ideation and problematic or binge drinking propensity from a shortened, 20-item brief instrument subset of the UPPS-P (Billieux et al., 2001). A resulting eight-item instrument was selected to be piloted for this research. Previous UPPS-P research indicated that five impulsivity facets were identified from the 59-items measure (Whiteside & Lynam, 2001). Four of the eight items selected were identified from facets that corresponded to problematic drinking: three items had a strong relationship to both problematic drinking and suicidal ideation, and the final item was identified only with suicidal ideation. Since the eight items chosen for inclusion in this research were selected to develop a brief impulsivity measure, Rasch analysis was used to determine item retention by examining dimensionality, fit, targeting, and scale use. Misfitting items were removed sequentially according to misfit, resulting in a four-item impulsivity measure.

The PHQ-4 was examined using hierarchical linear regression, with demographics and the impulsivity measure as independent variables. The results of this analysis showed a significant, positive contribution to explaining the variance at the second level due to the new impulsivity measure. This indicates that the addition of the impulsivity measure can aid in identifying depression that the PHQ-4 alone might not detect. The use of the four-item impulsivity measure with the PHQ-4 is an important outcome of this research, since the development of brief screening tools has been limited, with only a few instruments available that have very narrow targeting. This study also explored items from the other UPPS-P, which were used in the scale, and identified several items that could be used to develop a brief depression scale in the future.

For the AUDIT-C measure, the PHQ-4 and impulsivity measures did not appear to significantly enhance the identification for excessive alcohol consumption or binge drinking. Since Rasch analysis indicated that the AUDIT-C could not be used as a scale, the three questions were examined individually. For all three items, demographic variables (gender, age, and ethnicity) explained a significant amount of variation in the measure; however, neither the PHQ-4 nor the four-item impulsivity measure contributed to substantially explaining further variation. While this was a disappointing result, it was not completely unexpected, since previous research on drinking behaviors has consistently shown that gender, age, and ethnicity are by far the mostly closely related factors for predicting problematic drinking and binge drinking behaviors (Bradley et al., 2007). While the four-item impulsivity measure did not contribute to the better identification of problematic binge drinking, the researcher explored the contribution of individual impulsivity items from the original 15-item measure that was piloted. Three UPPS-P items were significant in the regression model, which is an important outcome of this study that confirms the need for more exploration between impulsivity and drinking behaviors. This is especially important given the need for better alcohol screening and the poor psychometric properties of the AUDIT-C for this population.

Finally, latent class analysis for this population revealed three classes that should help when targeting interventions for depression and identifying alcohol misuse. The groups themselves were not surprising, given the body of previous research (NIAA, n.d.;

Zakletskaia, Wilson, & Fleming, 2011), but rather confirms findings from the literature and contributes to identifying groups for prevention and education efforts.

In summary, the outcome of this research project was somewhat unexpected, but is important for guiding further research. This study showed a clear, but uneven relationship based on brief screening measures between impulsivity and both depression and alcohol misuse. For the depression scale, the impulsivity questions improved the explanation of the PHQ-4 variance more than expected, and produced an interesting result. The identification of depression is a challenge with a brief screening instrument due to targeting issues caused by the low occurrence of moderate to severe depression in the population and the prevalence of the generally low scores obtained by existing measures. Since this study was focused on choosing impulsivity questions to help identify excessive drinking behaviors, only a few of the impulsivity questions selected previously demonstrated a relationship with depression. However, the impulsivity questions demonstrated a significant contribution to depression prediction, and explained a significant amount of variance in the measure. The result of this study suggests a new direction for brief screening measures to improve the PHQ-4 and other brief screening instruments (such as the PHQ-9) in the future, with a broader set of items that have demonstrated a comorbidity with depression, rather than only focusing on the clinical definition of depression.

An equally interesting result was the association between the impulsivity items and the AUDIT-C measure. Prior research showed a strong relationship between problematic drinking and several impulsivity facets identified by the UPPS-P instrument. This study found similar results. Impulsivity items from the UPPS-P for this study did not add in a meaningful way to the overall score more than explained by basic demographics (gender, age, and ethnicity). In retrospect, this might be expected, as decades of research consistently linked demographic characteristics to problematic drinking. The lack of a relationship of the impulsivity questions casts further doubt on the ability to place reliance on the AUDIT-C as a useful tool for use in identifying problematic drinking, including binge drinking and drinking patterns that are life-long in nature.

The AUDIT-C demonstrated psychometric issues as a base, brief measure, first with the Rasch analysis, which demonstrated DIF. Also, in the hierarchical regression, overall variance explained was low. The highest variance explained by the regression model only explained less than 20% of the variance (AUDIT-C question 3). This indicates that there may be other factors that need to be included to improve this screening measure. Taken together, the research indicates that the AUDIT-C was of limited use in this population. Future research in developing a brief screening measure for problematic alcohol consumptions should be broader than simply asking “Do you drink?”; “Do you drink a lot?”; and “Do you often drink a lot when you drink a lot?”

Perhaps the most provoking outcome of this research is the indication that asking specific, clinically defined questions that are diagnostic in nature may not function adequately for brief screening measures. Simply asking fewer questions from a larger measure seems to result in inconsistent results when targeting, scale fit and use, item functioning, and contribution to variance explained are considered.

Perhaps a more effective method to develop brief screening instruments would be to consider the underlying issue more holistically, rather than approaching it using narrowly defined clinical guidelines. Including items that were not diagnostically driven, such as the impulsivity items to the PHQ-4 or the AUDIT-C, produced new ideas for future research. The addition of a wider range of questions identified by a broader body of research, such as family history or previous diagnoses, to depression or alcohol brief measures could result in both sounder brief screening instruments that result in better early identification so that early interventions may be pursued.

Optimally, development of a new instrument that would be more effective in screening for alcohol misuse and depression would include multiple research sites and begin with a pilot instrument that initially would have more items. The items would be selected from both clinical diagnostic criteria as well as from factors that research has shown to occur comorbidity, such as family history, previous issues with the underlying trait, as well as, selected demographical questions. Having a larger number of items with wider responses would give the researcher more ability to develop a truly effective screening instrument. Analytical procedures would be used to evaluate items fit, dimensionality, validity, and reliability. A key focus of these procedures would be to ensure that response scoring would be standardized to ensure better item difficulty, as well as ease of clinical interpretation. Using this method, instruments could be developed that would be psychometrically sound, explain a higher amount of what is now currently unexplained variance, and be diagnostically effective in brief screening settings.

Limitations

The limitations identified in this study included non-probability sampling, imbalances in demographic category choices, and survey design issues.

With the sampling strategy outside the control of the researcher, the generalizability of the results was a concern since randomization was not truly in place.

However, because the survey was turned on and off at random times during the survey period depending on a number of unrelated factors, when compared to previous visit statistics, the data appeared to be more random than convenient in nature. Administration of this instrument was standardized in the medical clinic to those with specified visit types. Improvements in sampling strategy, such as having the survey on at all times and for a longer period, would achieve a more robust sample.

The demographic groups used as categories for age and ethnicity could introduce bias into the models. Based on the research reviewed, there tends to be a delineation concerning the drinking patterns and levels of depression between undergraduate and graduate student. However, the cut-off age for these two groups (22 years and under, and 23 years and older) was the best estimate available based on the available enrollment and graduation information of the institution. While the ethnicity selector (white and non-white) was a natural decision, the population was a heavily weighted white population. Ethnicity differences should be inferred with this in mind. A larger sample would help the ability to generalize to a non-white demographic. Likewise, a larger sample of older students would have been helpful for a more robust generalization to older populations.

The design of the study and its use of an electronic entry format that limited the total amount of questions also possibly limited the research outcomes. If there had been the possibility of having longer scales by adding additional alcohol, depression, and impulsivity questions, there would have been more potential to have higher person separation. Essentially, the study attempted to generate a short scale out of other short scales, which might not have resulted in the best research outcome. The opportunity to use more questions from the different impulsivity facets might have increased the quality of the measure.

The nature of the questions, specifically in regard to alcohol consumption, may have resulted in the under-reporting of consumption and binge drinking numbers. There is a potential to not be truthful about alcohol consumption, especially when there is a perception that such information might be stored in a medical record. If the questionnaire was thought to be truly anonymous, there might have been more separation, as individuals might have reported higher drinking behavior that is more consistent with other reported consumption measures.

Recommendations for Further Study

The purpose of this research study was to develop and investigate the psychometric properties of a new brief screening measure for depression and problematic drinking, using impulsivity questions to enhance detection in the context of a brief screen. With this as a benchmark, the study succeeded in its purpose, but failed in that the resulting brief screen had inadequate reliability for each measure. However, important contributions to research in this area were uncovered that will aid in future research.

As noted earlier, the issue of person separation on a brief scale is an issue that should be explored further. The researcher did not find any IRT analyses on either of the brief screening instruments, the PHQ-4, the AUDIT-C, or the UPPS-P. While all of these scales have been used in research, and have different types of reported validity, the results of this study indicate that this is an area in need of further research. Part of the issue is the nature of a brief screening measure verses a diagnostic measure. The purpose of a brief screening measure is to identify a problem that may not be discovered, even simple questions like “Are you depressed?” are not asked. Once the problem is indicated, further diagnostic screening can be employed. However, it seems that with many brief screenings, the scale of development stops when respondents answer “Yes.”

The researcher believes that an “acceptable” balance can be achieved for brief screening instruments by balancing good scale development techniques with robust validity testing. One way to think of an acceptable balance for a brief screen would be to get enough information from the respondent so that the person administering the brief screen is unlikely to receive a false positive, and yet also minimizes the false negative and, therefore balances sensitivity and specificity. This is the true challenge of many of brief screens currently being applied for prevention and detection in primary medical care settings. Individuals are being identified for further assessment because they are episodically depressed due to an illness, or because they under-report, are missed by the brief screening instrument, and go undetected. As prevention and screening efforts for health issues become more widespread, effective use of brief screening instruments will enhance the efficient use of a clinician’s time, and can help maximize prevention efforts. That said, brief screening tools should be rigorously tested before they are widely adopted and before the conundrum of reliability versus utility understood.

The AUDIT-C measure in this study appears to be of questionable value as an instrument used to measure alcohol consumption in this population. Rasch analysis found differential item functioning for both gender and age. Further research with a broader population and through examining actual consumption to reported consumption should be pursued. With the prevalence of excessive consumption in the age group, it was disappointing to find that it did not perform well analytically as a measure. Given the widespread use of the AUDIT-C, and the health risks associated with excessive consumption, this should be a research priority.

Finally, the concept of impulsivity as an enhancement for a brief screening format is an open question. For the four-item impulsivity measure developed in this research, there was potential value, as there was for the PHQ-4 as a screening measure for depression. Further exploration by the researcher using individual items indicated that if the correct question was used, the impulsivity question may potentially contribute significantly to a scale that is focused in another area. Creative use of other questions (such as impulsivity) related to the underlying screening construct (such as depression or alcohol consumption) could be a way to achieve the person separation needed to make a brief screen more analytically robust. This, combined with effective targeting, could help continue improving models to help achieve effective brief screening instruments for use in clinical settings.

REFERENCES

Aalto, M., Alho, H., Halme, J. T., & Seppä, K. (2009). AUDIT and its abbreviated versions in detecting heavy and binge drinking in a general population survey. *Drug and Alcohol Dependence, 103*(1-2), 25-29. doi:10.1016/j.drugalcdep.2009.02.013

American College Health Association. *National College Health Assessment II: Reference group executive summary spring 2013*. Hanover, MD: American College Health Association.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5*. Washington, DC: American Psychiatric Association.

Arce, E., & Santisteban, C. (2006). Impulsivity: A review. *Psicothema, 18*(2), 213-220.

Babor, T. F., Higgins-Biddle, J. C., Saunders, J. B., & Monteiro, M. G. (2001). *The alcohol use disorders identification test: Guidelines for use in primary care* (2nd ed.). Geneva, Switzerland: World Health Organization.

Barratt, E. S. (1959). Anxiety and impulsiveness related to psychomotor efficiency. *Perceptual and Motor Skills, 9*, 191-198.

Barrios, L. C., Everett, S. A., Simon, T. R., & Brener, N. D. (2000). Suicide ideation among U.S. college students: Associations with other injury risk behaviors. *Journal of American College Health*, *48*(5), 229-233. doi:10.1080/07448480009599309

Beich, A., Thorsen, T., & Rollnick, S. (2003). Screening in brief intervention trials targeting excessive drinkers in general practice: Systematic review and meta-analysis. *British Medical Journal, 327*(7414), 536-542.

Bewick, B. M., Trusler, K., Barkham, M., Hill, A. J., Cahill, J., & Mulhern, B. (2008). The effectiveness of web-based interventions designed to decrease alcohol consumption: A systematic review. *Preventative Medicine, 47*(1), 17-26. doi:10.1016/j.ypmed.2008.01.005

Billieux, J., Rochat, L., Ceschi, G., Carré, A., Offerlin-Meyer, I., Defeldre, A. C., Van der Linden, M. (2012). Validation of a short French version of the UPPS-P impulsive behavior scale. *Comprehensive Psychiatry, 53*(5), 609-615. doi:10.1016/j.comppsych.2011.09.001

Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum.

Bradley, K. A., DeBenedetti, A. F., Volk, R. J., Williams, E. C., Frank, D., & Kivlahan, D. R. (2007). AUDIT-C as a brief screen for alcohol misuse in primary care. *Alcoholism, Clinical and Experimental Research, 31*(7), 1208-1217. doi:10.1111/j.1530-0277.2007.00403.x

Bush, K., Kivlahan, D. R., McDonell, M. B., Fihn, S. D., & Bradley, K. A. (1998). The AUDIT alcohol consumption questions (AUDIT-C): An effective brief screening test for problem drinking. *Archives of Internal Medicine, 158*(16), 1789.

Centers for Disease Control and Prevention. (2008). *Suicides Due to Alcohol and/or Drug Overdose: A Data Brief from the National Violent Death Reporting System*. Retrieved from http://www.cdc.gov/ViolencePrevention/pdf/NVDRS_Data_Brief-a.pdf

Centers for Disease Control and Prevention. (2011). An estimated 1 in 10 U.S. adults report depression. Retrieved from http://www.cdc.gov/features/dsdepression

Centers for Disease Control and Prevention. (2013). National ambulatory medical care survey: 2010 summary tables. Retrieved from: http://www.cdc.gov/nchs/data/ahcd/namcs_summary/2010_namcs_web_tables.pd f

Child, D. (2006). *The essentials of factor analysis*. London, UK: Continuum.

Clark, S. L. (2010). *Mixture modeling with behavioral data*. (Doctoral dissertation.). Retrieved from ProQuest.

Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis*. Hoboken, NJ: John Wiley & Sons, Inc.

Corruble, E., Danny, C., D., & Guelfi, G. (1999). Impulsivity: A relevant dimension in depression regarding suicide attempts? *Journal of Affective Disorders, 53*(3), 211-215. doi:10.1016/S0165-0327(98)00130-X

Coskunpinar, A., Dir, A. L., & Cyders, M. A. (2013). Multidimensionality in impulsivity and alcohol use: A meta-analysis using the UPPS model of impulsivity. *Alcoholism, Clinical and Experimental Research, 37*(9), 1441-1450. doi:10.1111/acer.12131

Cyders, M. A., Smith, G. T., Spillane, N. S., Fischer, S., Annus, A. M., & Peterson, C. (2007). Integration of impulsivity and positive mood to predict risky behavior: Development and validation of a measure of positive urgency. *Psychological Assessment, 19*(1), 107-118. doi:10.1037/1040-3590.19.1.107

DeVellis, R. F. (2003). *Scale development: Theory and applications*. Thousand Oaks, CA: Sage Publications, Inc.

Dawson, D. A., Smith, S. M., Saha, T. D., Rubinsky, A. D., & Grant, B. F. (2012). Comparative performance of the AUDIT-C in screening for DSM-IV and DSM-5 alcohol use disorders. *Drug and Alcohol Dependence*, *126*(3), 384-388.

Elliott, J. C., Carey, K. B., & Bolles, J. R. (2008). Computer-based interventions for college drinking: A qualitative review. *Addictive Behaviors, 33*(8), 994-1005. doi:10.1016/j.addbeh.2008.03.006

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272. doi:10.1037/1082-989X.4.3.27

Gliner, J. A., Morgan, G. A., & Leech, N. L. (2009). *Research methods in applied setttings: An integrated approach to design and analysis*. New York, NY: Routledge/Psychology Press.

Gonzalez, V. M., & Hewell, V. M. (2012b). Suicidal ideation and drinking to cope among college binge drinkers. *Addictive Behaviors, 37*(8), 994-997. doi:10.1016/j.addbeh.2012.03.027

Gorsuch, R. L. (1983). *Factor analysis*. Hillsdale, NJ: Lawrence Erlbaum & Associates. doi:113492078

Graham, A., Goss, C., Xu, S., Magid, D. J., & DiGuiseppi, C. (2007). Effect of using different modes to administer the AUDIT-C on identification of hazardous drinking and acquiescence to trial participation among injured patients. *Alcohol and Alcoholism*, *42*(5), 423-429. doi:10.1093/alcalc/agl123

Hill, R. M., Pettit, J. W., Green, K. L., Morgan, S. T., & Schatte, D. J. (2012). Precipitating events in adolescent suicidal crises: Exploring stress-reactive and non-reactive risk profiles. *Suicide Life Threat Behaviors, 42*(1), 11-21. doi:10.1111/j.1943-278X.2011.00067.x

Hyman, Z. (2006). Brief interventions for high-risk drinkers. *Journal of Clinical Nursing, 15*(11), 1383-1396. doi:10.1111/j.1365-2702.2006.01458.x

Jonas, D. E., Garbutt, J. C., Amick, H. R., Brown, J. M., Brownley, K. A., Council, C. L., Harris, R. P. (2012). Behavioral counseling after screening for alcohol misuse in primary care: A systematic review and meta-analysis for the U.S. Preventive services task force. *Annals of Internal Medicine, 157*(9), 645-654. doi:10.7326/0003-4819-157-9-201211060-00544

Kahan, M., Wilson, L., & Becker, L. (1995). Effectiveness of physician-based interventions with problem drinkers: A review. *Canadian Medical Association Journal, 152*(6), 851-859.

Kaner, E. F., Heather, N., Brodie, J., Lock, C. A., & McAvoy, B. R. (2001). Patient and practitioner characteristics predict brief alcohol intervention in primary care. *British Journal of General Practice, 51*(471), 822-827.

Koller, G., Preuss, U. W., Bottlender, M., Wenzel, K., & Soyka, M. (2002). Impulsivity and aggression as predictors of suicide attempts in alcoholics. *European Archives of Psychiatry and Clinical Neuroscience, 252*(4), 155-160. doi:10.1007/s00406-002-0362-9

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9. *Journal of General Internal Medicine*, *16*(9), 606-613.

Kroenke, K., Spitzer, R. L., Williams, J. B., & Löwe, B. (2010). The patient health questionnaire somatic, anxiety, and depressive symptom scales: A systematic review. *General Hospital Psychiatry*, *32*(4), 345-59. doi:10.1016/j.genhosppsych.2010.03.00

Kulesza, M., Apperson, M., Larimer, M. E., & Copeland, A. L. (2010). Brief alcohol intervention for college drinkers: How brief is? *Addictive Behaviors, 35*(7), 730-733. doi:10.1016/j.addbeh.2010.03.011

Lejuez, C. W., Magidson, J. F., Mitchell, S. H., Sinha, R., Stevens, M. C., & de Wit, H. (2010). Behavioral and biological indicators of impulsivity in the development of alcohol use, problems, and disorders. *Alcoholism, Clinical and Experimental Research*, *34*(8), 1334-1345. doi:10.1111/j.1530-0277.2010.01217.

Linacre, J. M. (2010). *A user's guide to WINSTEPS MINISTEP: Rasch-model computer programs* (Version 3.70.0) [computer software]. Chicago, IL: John M. Linacre.

Löwe, B., Spitzer, R. L., Williams, J. B., Mussell, M., Schellberg, D., & Kroenke, K. (2008). Depression, anxiety and somatization in primary care: Syndrome overlap and functional impairment. *General Hospital Psychiatry*, *30*(3), 191-199. doi:10.1016/j.genhosppsych.2008.01.00

Löwe, B., Wahl, I., Rose, M., Spitzer, C., Glaesmer, H., Wingenfeld, K., Brähler, E. (2010). A 4-item measure of depression and anxiety: Validation and standardization of the patient health questionnaire-4 (PHQ-4) in the general population. *Journal of Affective Disorders*, *122*(1-2), 86-95. doi:10.1016/j.jad.2009.06.019

Maddock, J. E., Laforge, R. G., Rossi, J. S., & O' Hare, T. (2001). The college alcohol problems scale. *Addictive Behaviors*, *26*(3), 385-398. doi:10.1016/S0306-4603(00)00116-7

MacCallum, R. C., Widaman, K. F., Preacher, K. J., & Hong, S. (2001). Sample size in factor analysis: The role of model error. *Multivariate Behavioral Research*, *36*(4), 611-637.

McCutcheon, A. L. (1987). *Latent class analysis*. Thousand Oaks, CA: Sage Publications.

Meneses-Gaya, C., Zuardi, A. W., Loureiro, S. R., Hallak, J. E. C., Trzesniak, C., de Azevedo Marques, J. M., Crippa, J. A. S. (2010). Is the full version of the AUDIT really necessary? Study of the validity and internal construct of its abbreviated versions. *Alcoholism, Clinical and Experimental Research, 34*(8), 1417-1424. doi:10.1111/j.1530-0277.2010.01225.x

Miller, E., Joseph, S., & Tudway, J. (2004). Assessing the component structure of four self-report measures of impulsivity. *Personality and Individual Differences*, *37*(2), 349-358.

Moeller, F. G., Barratt, E. S., Dougherty, D. M., Schmitz, J. M., & Swann, A. C. (2001). Psychiatric aspects of impulsivity. *American Journal of Psychiatry*, *158*(11), 1783-1793.

Muthén, L. K., & Muthén, B. O. (2012a). *Mplus* (Version 7) [computer software]. Los Angeles, CA: Muthén & Muthén.

Nagin, D. S. (2005). *Group-based modeling of development.* Cambridge, MA: Harvard University Press.

National Institute on Alcohol Abuse and Alcoholism (NIAAA). (n.d.). Moderate and binge drinking. Retrieved from http://www.niaaa.nih.gov/alcohol-your-health/overview-alcohol-consumption/moderate-binge-drinking

National Institute on Alcohol Abuse and Alcoholism (NIAA). (n.d.). College drinking. Retrieved from: http://www.niaaa.nih.gov/alcohol-health/special-populations-co-occurring-disorders/college-drinking

Neuhaus, V., & Ring, D. C. (2013). Latent class analysis. *The Journal of Hand Surgery*, *38*(5), 1018-1020. doi:10.1016/j.jhsa.2013.01.024

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. New York: McGraw-Hill.

Oöpik, P., Aluoja, A., Kalda, R., & Maaroos, H. I. (2006). Screening for depression in primary care. *Family Practice*, *23*(6), 693-698. doi:10.1093/fampra/cml05

Patient Health Questionnaire (PHQ) Screeners. Free Download. (n.d.). [Web page]. Retrieved from www.phqscreeners.com: http://www.phqscreeners.com/overview.aspx?Screener=01_PH

Patton, J. H., & Stanford, M. S. (1995). Factor structure of the Barratt impulsiveness scale. *Journal of Clinical Psychology, 51*(6), 768-774.

Pyne, J. M., Rost, K. M., Zhang, M., Williams, D. K., Smith, J., & Fortney, J. (2003). Cost-effectiveness of a primary care depression intervention. *Journal of General Internal Medicine, 18*(6), 432-441.

Rost, K., Nutting, P., Smith, J. L., Elliott, C. E., & Dickinson, M. (2002). Managing depression as a chronic disease: A randomised trial of ongoing treatment in primary care. *British Medical Journal, 325*(7370), 934.

Saitz, R., Helmuth, E. D., Aromaa, S. E., Guard, A., Belanger, M., & Rosenbloom, D. L. (2004). Web-based screening and brief intervention for the spectrum of alcohol problems. *Preventative Medicine, 39*(5), 969-975. doi:10.1016/j.ypmed.2004.04.011

Shin, S. H., Hong, H. G., & Jeon, S. M. (2012). Personality and alcohol use: The role of impulsivity. *Addictive Behaviors, 37*(1), 102-107. doi:10.1016/j.addbeh.2011.09.006

Simon, S., Steward, K., Kloc, M., Williams, T. V., & Wilmoth, M. C. (2013). The reliability of a mental health screening and assessment instrument designed for deployed members of the armed forces. Proceedingss of the Academy Health Research Network Conference.

Smith, P. C., Schmidt, S. M., Allensworth-Davies, D., & Saitz, R. (2009). Primary care validation of a single-question alcohol screening test. *Journal of General Internal Medicine, 24*(7), 783-788. doi:10.1007/s11606-009-0928-6

Stanford, M. S., Mathias, C. W., Dougherty, D. M., Lake, S. L., Anderson, N. E., & Patton, J. H. (2009). Fifty years of the Barratt impulsiveness scale: An update and review. *Personality and Individual Differences*, *47*(5), 385-395.

Stark, S., Chernyshenko, S., Chuah, D., Lee, W., & Wadlington, P. (2001). Selecting a dichtomous IRT model. [On-line tutorial]. Retrieved from: http://work.psych.uiuc.edu/irt/modeling_dich1.asp

Strosahl, K. (1996). Confessions of a behavior therapist in primary care: The odyssey and the ecstasy. *Cognitive and Behavioral Practice*, *3*(1), 1-28.

Substance Abuse and Mental Health Services. (2008). *Testing the difference between the highest and lowest prevalence rates in substate regions within each state based on data collected from the 2004-2006 National Surveys on Drug Use and Health*. Retrieved from: http://www.samhsa.gov/data/substate2k8/scales/scales.pdf

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Pearson Education, Inc.

Verdejo-García, A., Lozano, Ó., Moya, M., Alcázar, M. Á., & Pérez-García, M. (2010). Psychometric properties of a spanish version of the UPPS–P impulsive behavior scale: Reliability, validity and association with trait and cognitive impulsivity. *Journal of Personality Assessment*, *92*(1), 70-77. doi:10.1080/00223890903382369

Vermunt, J. K., & Magidson, J. (2002). Latent class cluster analysis. Retrieved from: http://www.statisticalinnovations.com/articles/Latclass.pdf

Walker, E. R., Engelhard, G., & Thompson, N. J. (2012). Using Rasch measurement theory to assess three depression scales among adults with epilepsy. *Seizure: The*

Journal of the British Epilepsy Association, 21(6), 437-443. doi:10.1016/j.seizure.2012.04.009

Wang, J., & Wang. X. (2012). Structural equation modeling applications using *MPLUS*. Chichester, West Sussex, UK: Wiley. doi:9781118356319

Whiteside, S. P., & Lynam, D. R. (2001). The five factor model and impulsivity: Using a structural model of personality to understand impulsivity. *Personality and Individual Differences, 30*(4), 669-689.

Whitlock, E. P., Polen, M. R., Green, C. A., Orleans, T., Klein, J., & U.S. Preventive Services Task Force. (2004). Behavioral counseling interventions in primary care to reduce risky/harmful alcohol use by adults: A summary of the evidence for the U.S. Preventive Services Task Force. *Annals of Internal Medicine*, *140*(7), 557-568.

Yu, C. H., Popp, S. O., DiGangi, S., & Jannasch-Pennell, A. (2007). Assessing unidimensionality: A comparison of Rasch modeling, parallel analysis, and TETRAD. *Practical Assessment, Research & Evaluation, 12*(14) 19 pages.

Zakletskaia, L., Wilson, E., & Fleming, M. F. (2011). Alcohol use in students seeking primary care treatment at university health services. *Journal of American College Health*, *59*(3), 217-223. doi:10.1080/07448481.2010.502413

APPENDIX A: PILOT SCALE

Over the last 2 weeks, how often have you been bothered by the following problems?

1. Feeling nervous, anxious or on edge
2. Not being able to stop or control worrying
3. Little interest or pleasure in doing things
4. Feeling down, depressed, or hopeless
5. How often do you have a drink containing alcohol?
6. How many drinks containing alcohol do you have on a typical day when you are drinking?
7. How often do you have six or more drinks on one occasion?
8. I usually think carefully before doing anything
9. I finish what I start
10. When I am really excited, I tend not to think on the consequences of my actions
11. I tend to act without thinking when I am really excited
12. When I am upset, I often act without thinking
13. I often make matters worse because I act without thinking when I am upset
14. I sometimes like doing things that are a bit frightening
15. I welcome new and exciting experiences and sensations, even if they are a little frightening and unconventional

Scoring

Questions one through four are the PHQ-4 instrument and are scored:

0 - Not at all
1- Several days
2 - More than half the days
3 - Nearly every day (4)

Questions five on the AUDIT-C instrument are scored:

0 – Never
1 - Monthly or less
2 - Two to four times a month
3 - Two to three times a week
4 - Four or more times a week

Questions six the AUDIT-C instrument is scored:

0 – 1 or 2
1 – 3 or 4
2 – 5 or 6
3 – 7 to 9
4 – 10 or more

Questions seven on the AUDIT-C instrument is scored:

0 – Never
1 – Less than monthly
2 - Monthly
3 - Weekly
4 – Daily or almost daily

Questions 8-15 are from the UPPS-P and the shortened UPPS-P scales and are scored:

1 - Agree Strongly
2 - Agree Some
3 - Disagree Some
4 - Disagree strongly

*The numbers behind the impulsivity questions are the item number on the full UPPS-P and the short UPPS-P instruments respectively. The italicized text indicates the impulsivity facet(s) and disorder previous research has shown they are intended to measure.

Question 8 - (48/1) *Premeditation (lack of): (suicide ideation - Bipolar Disorder*
Question 9 - (27/8) *Perseverance (lack of): (suicide ideation - alcohol use)*
Question 10 - (R 50/2) *Positive Urgency: (AUDs, binge drinking)*
Question 11 - (R 53/15) *Positive Urgency: (AUDs, binge drinking)*
Question 12 - (R 29/4) *Negative Urgency: (drinking to cope, drinking problems, suicidal ideation)*
Question 13 - (R 44/12) *Negative Urgency: (drinking to cope, drinking problems, suicidal ideation)*
Question 14 – (R 41/3) *Sensation Seeking: (AUDs, Frequency of drinking, binge drinking)*
Question 15 - (R 31/18) *Sensation Seeking: (AUDs, Frequency of drinking, binge drinking)*

The “R” signals reverse scoring. Italicized indicates the UPPS-P facet and the correlation to a behavioral trait.

APPENDIX B: PHQ-4 QUESTIONNAIRE

PHQ-4 Questions: Scored 0-3

Over the last 2 weeks, how often have you been bothered by the following problems?

1. Feeling nervous, anxious or on edge

2. Not being able to stop or control worrying

3. Little interest or pleasure in doing things

4. Feeling down, depressed, or hopeless

Scoring:

Not at all (0), Several days (1), More than half the days (2), Nearly every day (3)

APPENDIX C: AUDIT-C SCREENING QUESTIONNAIRE

1. How often do you have a drink containing alcohol?
2. How many drinks containing alcohol do you have on a typical day when you are drinking?
3. How often do you have six or more drinks on one occasion?

Scoring:

Never 0, Less than monthly 1, Monthly 2, Weekly 3, Daily or almost daily 4

Scoring: Sum of the scores for the three questions

APPENDIX D: POPULATION DEMOGRAPHICS

Age

Mean:	22.38
Median:	21
Standard Deviation:	4.298
Skewness:	1.695
Kertosis:	2.968
Range:	18-44

Age Groups for Rasch:

1) 18-22 (undergraduates)
2) 23 and older (graduates)

Age Frequencies

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	18.0	32	6.5	6.5	6.5
	19.0	95	19.3	19.3	25.9
	20.0	88	17.9	17.9	43.8
	21.0	72	14.7	14.7	58.5
	22.0	54	11.0	11.0	69.5
	23.0	25	5.1	5.1	74.5
	24.0	19	3.9	3.9	78.4
	25.0	7	1.4	1.4	79.8
	26.0	20	4.1	4.1	83.9
	27.0	13	2.6	2.6	86.6
	28.0	11	2.2	2.2	88.8
	29.0	12	2.4	2.4	91.2
	30.0	11	2.2	2.2	93.5
	31.0	8	1.6	1.6	95.1
	32.0	4	.8	.8	95.9
	33.0	6	1.2	1.2	97.1
	34.0	4	.8	.8	98.0
	35.0	4	.8	.8	98.8
	36.0	1	.2	.2	99.0
	38.0	3	.6	.6	99.6
	41.0	1	.2	.2	99.8
	44.0	1	.2	.2	100.0
	Total	491	100.0	100.0	

Gender	
Female:	64.8%
Male:	35.2%
Ethnicity	
White:	79.4%
Non-White	20.6%

APPENDIX E: ITEM RESPONSE BY ITEM

PHQ-4 Question 1

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	295	60.1	60.1	60.1
	1	145	29.5	29.5	89.6
	2	31	6.3	6.3	95.9
	3	20	4.1	4.1	100.0
	Total	491	100.0	100.0	

PHQ-4 Question 2

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	333	67.8	67.8	67.8
	1	127	25.9	25.9	93.7
	2	18	3.7	3.7	97.4
	3	13	2.6	2.6	100.0
	Total	491	100.0	100.0	

PHQ-4 Question 3

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	389	79.2	79.2	79.2
	1	79	16.1	16.1	95.3
	2	15	3.1	3.1	98.4
	3	8	1.6	1.6	100.0
	Total	491	100.0	100.0	

PHQ-4 Question 4

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	399	81.3	81.3	81.3
	1	74	15.1	15.1	96.3
	2	15	3.1	3.1	99.4
	3	3	.6	.6	100.0
	Total	491	100.0	100.0	

AUDIT-C Question 1

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	89	18.1	18.1	18.1
	1	83	16.9	16.9	35.0
	2	152	31.0	31.0	66.0
	3	143	29.1	29.1	95.1
	4	24	4.9	4.9	100.0
	Total	491	100.0	100.0	

AUDICT-C Question 2

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	285	58.0	58.0	58.0
	1	153	31.2	31.2	89.2
	2	37	7.5	7.5	96.7
	3	14	2.9	2.9	99.6
	4	2	.4	.4	100.0
	Total	491	100.0	100.0	

AUDIT-C Question 3

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	221	45.0	45.0	45.0
	1	163	33.2	33.2	78.2
	2	67	13.6	13.6	91.9
	3	40	8.1	8.1	100.0
	Total	491	100.0	100.0	

UPPS-P Question 48

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	336	68.4	68.4	68.4
	2	145	29.5	29.5	98.0
	3	9	1.8	1.8	99.8
	4	1	.2	.2	100.0
	Total	491	100.0	100.0	

UPPS-P Question 27

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	33	6.7	6.7	6.7
	2	106	21.6	21.6	28.3
	3	166	33.8	33.8	62.1
	4	186	37.9	37.9	100.0
	Total	491	100.0	100.0	

UPPS-P Question 50

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	43	8.8	8.8	8.8
	2	215	43.8	43.8	52.5
	3	124	25.3	25.3	77.8
	4	109	22.2	22.2	100.0
	Total	491	100.0	100.0	

UPPS-P Question 53

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	13	2.6	2.6	2.6
	2	76	15.5	15.5	18.1
	3	181	36.9	36.9	55.0
	4	221	45.0	45.0	100.0
	Total	491	100.0	100.0	

UPPS-P Question 29

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	274	55.8	55.8	55.8
	2	185	37.7	37.7	93.5
	3	23	4.7	4.7	98.2
	4	9	1.8	1.8	100.0
	Total	491	100.0	100.0	

UPPS-P Question 44

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	16	3.3	3.3	3.3
	2	77	15.7	15.7	18.9
	3	142	28.9	28.9	47.9
	4	256	52.1	52.1	100.0
	Total	491	100.0	100.0	

UPPS-P Question 31

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	142	28.9	28.9	28.9
	2	258	52.5	52.5	81.5
	3	64	13.0	13.0	94.5
	4	27	5.5	5.5	100.0
	Total	491	100.0	100.0	

UPPS-P Question 41

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	15	3.1	3.1	3.1
	2	105	21.4	21.4	24.4
	3	163	33.2	33.2	57.6
	4	208	42.4	42.4	100.0
	Total	491	100.0	100.0	

APPENDIX F: MPLUS INPUT FILE EXAMPLE

```
Mplus VERSION 7.11
MUTHEN & MUTHEN
09/19/2014   7:42 PM

INPUT INSTRUCTIONS

  TITLE: Wera Dissertation LCA

  DATA:
    FILE IS "E:\Wera_Dissertation_mPlus.dat";

   VARIABLE:
     NAMES ARE gender age ethnic phq_4 auditc1 auditc2 Auditc3
impul;
      USEVARIABLES ARE gender age ethnic phq_4 auditc1 auditc2
Auditc3 impul;
      CATEGORICAL ARE gender-ethnic;
      CLASSES = c(4);

    ANALYSIS:
      TYPE IS MIXTURE;

        LOGHIGH = +15;
      LOGLOW = -15;
      UCELLSIZE = 0.01;
      ESTIMATOR IS MLR;
      LOGCRITERION = 0.0000001;
      ITERATIONS = 1000;
      CONVERGENCE = 0.000001;
      MITERATIONS = 500;
      MCONVERGENCE = 0.000001;
      MIXC = ITERATIONS;
      MCITERATIONS = 2;
      MIXU = ITERATIONS;
      MUITERATIONS = 2;

    PLOT:
        type is plot3;
          series is  phq_4 (1) auditc1 (2) auditc2 (3) Auditc3
(4) impul (5);
```

Figure 8. Example of a typical *Mplus* input file.

APPENDIX G: ABBREVIATIONS AND ACRONYMS

ABIC	Adjusted Bayesian Information Criterion
AIC	Akaike’s Information Criterion
ALMR LR	Adjusted Lo-Mendell-Rubin Likelihood Ratio
ANOVA	Analysis of Variance
AVE	Average Variance Extracted
BIC	Bayesian Information Criterion
BLRT	Bootstrap Likelihood Ratio Test
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
CTT	Classical Test Theory
DIF	Differential Item Functioning
EFA	Exploratory Factor Analysis
EM	Expectation-Maximization
EPC	Expected Parameter Change
ICC	Item Characteristic Curves
IRB	Institutional Review Board
IRT	Item Response Theory
LCA	Latent Class Analysis
LRM LR	Lo-Mendell-Rubin Likelihood Ratio
MI	Modification Indices

MIRT	Multidimensional Item Response Theory
ML	Maximum Likelihood
MRCMLM	Multidimensional Random Coefficients Multinomial Logit Model
ORF	Option Response Function
RCMLM	Random Coefficients Multinomial Logit Model
RMSEA	Root Mean Square Error of Approximation
RSM	Rating-Scale Model
SEM	Structural Equation Modeling
TLI	Tucker-Lewis Fit Index
ULI	Unit Loading Identification
VIF	Variance Inflation Factor
WLS	Weighted Least Squares
WLSMV	Means and Variances Corrected Diagonally Weighted Least Squares
WRMR	Weighted Root Mean Square Residual

APPENDIX H: UNIVERSTIY OF DENVER INSTITUTIONAL REVIEW BOARD

(IRB) APPROVAL

UNIVERSITY of DENVER
RESEARCH & SPONSORED PROGRAMS

DATE:	April 25, 2014
TO:	Chris Wera, MACC
FROM:	University of Denver (DU) IRB
PROJECT TITLE:	[578605-2] Modification of PHQ-9 intake questionnaire at the DU Health and Counseling Center
SUBMISSION TYPE:	Response/Follow-Up
ACTION:	APPROVED
APPROVAL DATE:	April 25, 2014
EXPIRATION DATE:	April 24, 2015
REVIEW TYPE:	Expedited Review
REVIEW CATEGORY:	Expedited review category # 7

Thank you for your submission of Response/Follow-Up materials for this project. The University of Denver (DU) IRB has APPROVED your submission. This approval is based on an appropriate risk/benefit ratio and a project design wherein the risks have been minimized. All research must be conducted in accordance with this approved submission.

This submission has received Expedited Review based on the applicable federal regulations.

Please remember that informed consent is a process beginning with a description of the project and insurance of participant understanding followed by a signed consent form. Informed consent must continue throughout the project via a dialogue between the researcher and research participant. Federal regulations require each participant receives a copy of the consent document.

Please note that any revision to previously approved materials must be approved by this office prior to initiation. Please use the appropriate revision forms for this procedure.

All UNANTICIPATED PROBLEMS involving risks to subjects or others and SERIOUS and UNEXPECTED adverse events must be reported promptly to this office. Please use the appropriate reporting forms for this procedure. All FDA and sponsor reporting requirements should also be followed.

All NON-COMPLIANCE issues or COMPLAINTS regarding this project must be reported promptly to this committee.

This project has been determined to be a Minimal Risk project. Based on the risks, this project requires continuing review by this committee on an annual basis. Please use the appropriate forms for this procedure. Your documentation for continuing review must be received with sufficient time for review and continued approval before the expiration date of April 24, 2015.

Please note that all research records must be retained for a minimum of three years after the completion of the project.